



Departamento de Medicina Preventiva y Salud
Pública y Microbiología

DESARROLLO Y APLICACIÓN DE MODELOS PRONÓSTICOS EN PATOLOGÍA LUMBAR

Tesis Doctoral

Ana Royuela Vicente

Madrid, 2014

Este trabajo ha sido financiado por el Consorcio de Investigación Biomédica de Epidemiología y Salud Pública (CIBERESP). Ana Royuela Vicente disfruta de un contrato como técnico investigador y ha recibido una beca para una estancia en el extranjero financiada por el CIBERESP.

*“Quien no haya experimentado la irresistible atracción de la ciencia,
no podrá comprender su tiranía.”*

Frankenstein. Mary Shelley

AGRADECIMIENTOS

Al finalizar un trabajo tan arduo como el desarrollo de una Tesis Doctoral, es inevitable que acuda a la conciencia que el llegar hasta aquí hubiera sido imposible sin la presencia de algunas personas. Personas que han facilitado bien desde un punto de vista profesional y/o personal que ahora pueda presentar esta Tesis. Por ello, es para mí un honor y un placer utilizar este espacio para expresarles mis agradecimientos:

A mis padres, que han hecho tanto en esta vida para que pudiera lograr mis sueños.

A Jaime, por su paciencia y comprensión. Por sus “primero la Tesis y luego ya veremos”.

A mis compañeros de la Unidad de Bioestadística, en especial al Dr. Víctor Abaira, mi director de Tesis. Habéis sido mis maestros, compañeros y amigos durante todos estos años, gracias por enseñarme tantas cosas.

A la Profesora Pilar Guallar-Castillón por su tutorización. Al Departamento de Medicina Preventiva, Salud Pública y Microbiología de la Facultad de Medicina de la Universidad Autónoma de Madrid, por su confianza y por brindarme la oportunidad de realizar y defender esta Tesis Doctoral.

Al Dr. Francisco Kovacs, por su apoyo y generosidad durante tantos años. Sin él, no habría sido posible.

A la Red Española de Investigadores en Dolencias de la Espalda (REIDE), por compartir la información clínica para la realización de esta Tesis.

Al Consorcio de Investigación Biomédica de Epidemiología y Salud Pública (CIBERESP). Gracias a su financiación he podido desarrollar esta Tesis Doctoral y tuve la oportunidad de visitar durante unos meses el *Department of Primary Care Health Sciences* de la Universidad de Oxford (Reino Unido).

Al grupo de amigos que surgió en la Facultad de Biología y con el tiempo ha ido incrementándose, por tan buenos momentos. Gracias especialmente a los ya Doctores, por sus ánimos y sabios consejos.

A todos aquellos familiares, amigos y compañeros del Hospital Ramón y Cajal: mi hermano Pablo, mi tía Ana, Ainhoa, Vega, Cris, Esther, Isa, Sandra, Irene, Marisa, Rober, Sebas, Dani, Mono,... no tengo espacio suficiente para nombraros pero habéis estado ahí, apoyándome para que esta Tesis saliera adelante. Os quiero.

ÍNDICE

	Pág.
RESUMEN	27
ABSTRACT	31
INTRODUCCIÓN	35
INTRODUCTION	47
HIPÓTESIS	57
OBJETIVOS	61
MATERIAL Y MÉTODOS Y RESULTADOS	
ARTÍCULO 1: <i>Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service</i>	
Justificación metodológica	67
Participantes y métodos	71
Resultados: Artículo publicado y aplicación web	76
ARTÍCULO 2: <i>Predicting outcomes of neuroreflexotherapy in patients with subacute or chronic neck or low back pain</i>	
Justificación metodológica	95
Participantes y métodos	100
Resultados: Artículo publicado y aplicación web	108
DISCUSIÓN	133
CONCLUSIONES	141
CONCLUSIONS	145
APÉNDICES	149
Apéndice I: Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. <i>Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain</i> . Spine (Phila Pa 1976). 2007 Dec 1;32(25):2915-20.	
Apéndice II: Reglas de Rubin	

Apéndice III: Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomás M, et al. *Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain*. BMC Musculoskelet Disord. 2008 Apr 10;9:43.

Apéndice IV: Kovacs FM, Bagó J, Royuela A, Seco J, Giménez S, Muriel A, et al. *Psychometric characteristics of the Spanish version of instruments to measure neck pain disability*. BMC Musculoskelet Disord. 2008 Apr 9;9:42.

Apéndice V: Royuela A, Abriara V. *Calibration plot joining Multiple Imputation Dataset*. Eur J Epidemiol (2013) 28: S1-S270

Apéndice VI: Eficiencia de la imputación múltiple

BIBLIOGRAFÍA **191**

Índice de Tablas y Figuras **199**

Lista de acrónimos

BMJ	British Medical Journal
CVRS	Calidad de Vida Relacionada con la Salud
FMI	Fracción de Información Perdida
GLM	Modelo Lineal Generalizado (Generalized Linear Model)
ISI	Institute for Scientific Information
JAMA	Journal of the American Medical Association
JCR	Journal Citation Report
MAR	Missing At Random
MCAR	Missing Completely At Random
MICE	Multiple Imputation by Chained Equations
MNAR	Missing Not At Random
NDI	Neck Disability Index
NEJM	The New England Journal of Medicine
NRT	Intervención Neurorreflejo terapéutica
OMC	Organización Médica Colegial
PIB	Producto Interior Bruto
RE	Eficiencia Relativa de la Imputación
REIDE	Red Española de Investigadores en Dolencias de la Espalda
ROC	Receiver Operating Characteristic
SNS	Sistema Nacional de Salud
VAS	Escala Visual Analógica (Visual Analog Scale)

En esta Tesis Doctoral se presentan dos artículos publicados en la revista *The Spine Journal*. Posee un Factor de Impacto igual a 3,220 puntos (Edición científica JCR 2012). Ocupa el cuarto puesto sobre un total de 65 revistas en la categoría *Orthopedics*, indexada por el *Institute for Scientific Information* (ISI), encontrándose por tanto, en su primer decil. Además, es considerada la principal revista científica en el campo del raquis y la revista oficial de la correspondiente Sociedad Norteamericana (*North American Spine Society* -NASS-).

Parte de los resultados han sido presentados en el Congreso Europeo de Epidemiología (EUROEPI), celebrado durante el mes de agosto de 2013 en la ciudad danesa de Aarhus. La presentación fue un póster titulado: "*Calibration plot joining multiple imputation datasets*".

Se ha desarrollado una aplicación web con los resultados de los modelos pronósticos que son producto de esta Tesis. Así, tanto los pacientes como profesionales interesados, pueden conocer la probabilidad estimada de mejoría de su dolor y discapacidad bajo determinadas circunstancias:

www.pronosticodolorlumbar.org

RESUMEN

Antecedentes: A pesar de las numerosas publicaciones sobre procedimientos para tratar el dolor lumbar inespecífico, no se conocen qué factores se asocian a una mejoría clínicamente relevante en relación a la intensidad de dolor, local e irradiado, y la discapacidad asociada al dolor lumbar.

En pacientes que padecen lumbalgia o cervicalgia inespecífica, intervenidos mediante neuroreflejo terapia (NRT), se desconoce qué factores se asocian a una mejoría clínicamente relevante en cuanto a la intensidad del dolor local, irradiado y discapacidad.

Métodos: Para conocer qué factores pronósticos se asocian a una mejoría clínicamente relevante en la intensidad de dolor lumbar e irradiado a la pierna y la discapacidad asociada, se estudiaron los pacientes incluidos en una cohorte prospectiva de casi 5.000 sujetos que acudieron por dolor lumbar inespecífico a su médico. Basalmente se recogieron variables clínicas y sociodemográficas. El seguimiento fue de tres 3 meses, momento en que se volvieron a recoger las variables relacionadas con la intensidad de dolor y la discapacidad.

Una segunda cohorte prospectiva compuesta por casi 9.000 pacientes con lumbalgia o cervicalgia inespecífica que recibieron una intervención NRT, facilitó los datos para conocer los factores pronósticos asociados a la mejoría clínicamente relevante en relación a la intensidad del dolor local e irradiado y discapacidad asociada, tres meses después de dicha intervención.

Partiendo de las asunciones necesarias, se llevaron a cabo imputaciones múltiples de datos faltantes. Se desarrollaron modelos pronósticos mediante regresiones logísticas múltiples. Para valorar el rendimiento de los modelos, se evaluó su calibración y discriminación, así como su validez interna mediante técnicas de *bootstrapping*.

Para facilitar la interpretación de los resultados de los modelos de regresión, se desarrollaron nomogramas y calculadoras web, facilitando la accesibilidad a estos resultados por parte del público no especializado.

Resultados: En la primera cohorte, los factores asociados a una mayor probabilidad de mejoría, comunes a la intensidad de dolor y la discapacidad fueron: no haber sido intervenido quirúrgicamente del segmento doloroso, valores basales altos en intensidad de dolor y discapacidad en sus respectivos modelos, pero menores en el

resto, y haber sido tratados con la intervención NRT. La calibración y discriminación de los tres modelos resultaron razonables.

En la segunda cohorte, las variables que resultaron tener un mayor valor en el pronóstico de los tres modelos fueron: niveles basales de intensidad de dolor, local e irradiado y discapacidad, duración del dolor, realización de una radiografía simple, haber sido intervenido quirúrgicamente del segmento doloroso y encontrarse inmerso en un proceso de reclamación laboral por dolor cervical o lumbar. En estos tres modelos, la discriminación osciló entre ligera y moderada y la calibración fue buena.

Para ambos estudios, se han desarrollado sus correspondientes nomogramas y una calculadora web que proporciona la probabilidad de mejoría del dolor local, dolor irradiado y discapacidad.

Conclusiones: El empleo de métodos estadísticos avanzados permite conocer de forma válida las variables asociadas al pronóstico del dolor lumbar y cervical en dos cohortes de pacientes.

El desarrollo de herramientas como nomogramas y calculadoras web, facilitan la interpretación y difusión de resultados en investigación.

ABSTRACT

Background: Despite numerous publications about procedures in order to treat low back pain, factors associated with a clinically relevant improvement of pain and disability are still unknown.

In patients with non-specific low back and neck pain who have been treated with neuroreflexotherapy (NRT), there is a lack of knowledge about factors associated with a clinically relevant improvement in their pain intensity and disability.

Methods: In order to know the associated factors with a clinically relevant improvement in the intensity of local and referred pain also disability, a prospective cohort with almost 5,000 subjects who attended their physician for non-specific low back pain were recruited. Sociodemographic and clinical baseline variables were recorded. The follow-up was for three months, when variables related with pain intensity and disability were recorded again.

The second prospective cohort is compound by almost 9,000 patients with non-specific low back and neck pain, who have been treated with NRT. Patients were recruited in order to know the associated prognostic factors with a clinically relevant improvement in the intensity of local and referred pain also disability, three months after the NRT intervention.

From some required assumptions, multiple imputations of missing data were performed. Prognostic models through multiple logistic regressions were developed. To assess the models performance, their calibration and discrimination were both evaluated, as well as the internal accuracy by means of *bootstrapping* techniques.

In order to ease the interpretation of regression models results, nomograms and web calculators were developed, making also easier the accessibility for non-specialized public to research results.

Results: In the first cohort, common factors associated with a higher probability of improvement in pain intensity and disability were as follows: not having undergone lumbar surgery, higher baseline scores for the corresponding variable, lower ones for the rest, and being treated with NRT. Calibration and discrimination for the three models were reasonable.

In the second cohort, variables which showed a higher predictive value among the three prognostic models were: baseline scores for pain and disability, pain duration,

having undergone X-ray, having undergone spine surgery, and receiving financial assistance for neck or low back pain. Discrimination of the three models ranged from slight to moderate, and calibration was good.

For both studies, their corresponding nomograms and web calculators providing the probabilities of improvement in pain intensity and disability were developed.

Conclusions: The use of advanced statistical methods allow to know accurately which variables are associated to prognosis of non-specific low back and neck pain throughout two different patient cohorts.

The development of tools as nomograms and web calculators, ease the interpretation also dissemination of research results.

INTRODUCCIÓN

Pronóstico significa anticipo, predicción o estimación de la probabilidad o riesgo de situaciones futuras. En medicina, más concretamente, el concepto de pronóstico se relaciona con la probabilidad o riesgo de un individuo de desarrollar un estado de salud determinado (un “outcome”) a lo largo de un tiempo específico, basándose tanto en su perfil clínico como no clínico.

Bajo este contexto, los conceptos factor de riesgo y factor pronóstico podrían parecer equivalentes, mas existen diferencias conceptuales entre ellos. El factor de riesgo puede estar presente en población sana y su presencia se asocia con un incremento en la probabilidad de padecer una enfermedad. Es la probabilidad de ocurrencia de un hecho epidemiológico. En cambio, un factor pronóstico, necesariamente parte de una situación de enfermedad y predice el curso clínico una vez que la enfermedad se ha presentado (1,2).

Dada la variabilidad presente entre pacientes, su etiología, condiciones de presentación de la enfermedad y tratamientos prescritos, un sólo factor pronóstico o predictor difícilmente proporcionará una adecuada estimación del pronóstico. Los estudios de pronóstico necesitan por tanto, utilizar varios predictores de cara a estimar la evolución de la enfermedad en un paciente. Es necesaria una aproximación multivariable, tanto en el diseño como en el análisis, para determinar qué predictores relevantes se asocian a un resultado, proporcionar probabilidades que combinen diferentes variables predictoras, o facilitar herramientas que estimen estas probabilidades. Estas herramientas se denominan modelos pronósticos (3,4).

En el diseño de un estudio que evalúe el pronóstico, la recogida de datos a lo largo del eje temporal es el punto de partida. Desde ahí, pueden diseñarse estudios de cohortes, casos y controles o ensayos clínicos, según los objetivos del investigador. Si el objetivo es observar correctamente la secuencia temporal entre exposición y desenlace, la “historia natural” de la enfermedad, estimar incidencias, tener un buen control sobre la selección de sujetos y mediciones, además de moverse en un escenario lo más parecido posible a la práctica clínica, el diseño ideal es una cohorte prospectiva.

Cabe destacar que, a pesar de la similitud en el diseño y análisis entre modelos pronósticos y estudios etiológicos, la predicción de resultados no es sinónimo de

justificar su causalidad. Con los estudios pronósticos no se pretende establecer la causalidad entre un determinado factor y un resultado, sino utilizar múltiples variables para poder predecir con precisión el riesgo de futuros eventos. Asociación no siempre implica causalidad. Los estimadores del riesgo utilizados en estudios de pronóstico (*odds ratio*, riesgo relativo o *hazard ratio*), no son necesariamente causales. Son utilizados para obtener una probabilidad de resultado en un individuo concreto. Como contraste, los estudios etiológicos se centran en estimar el riesgo de ocurrencia de un resultado en presencia de un factor causal relativo al riesgo de ocurrencia en ausencia de ese factor (3,5).

Para llevar a cabo una investigación sobre pronóstico, el punto de partida debe ser el desarrollo del modelo, después validar su rendimiento y finalmente, estudiar su impacto clínico. Los modelos de regresión multivariante permiten estimar, a partir de una serie de variables, la probabilidad que un paciente tiene de mejorar en su enfermedad en un tiempo determinado (6). Según las características de la variable dependiente, el modelo puede ser lineal o logístico, y si se quiere tener en cuenta el tiempo hasta que sucede el evento, el modelo a elegir será un modelo de riesgo proporcional o de Cox.

Si el objetivo es predecir qué variables se asocian a una variable binomial, el modelo adecuado es regresión logística. Se encuadra dentro del Modelo Lineal Generalizado (GLM), y utiliza como función de enlace la función *logit*. La probabilidad que describe el resultado se modela como una función de las variables explicativas, utilizando una función logística.

Cualquier estudio de investigación, ya sea observacional o experimental, está expuesto a la pérdida de datos. Esto conlleva una pérdida de potencia en el análisis de los resultados, puesto que la tendencia tradicional ha sido la de prescindir de aquellas observaciones con falta de datos. Además, si la razón por la que estos datos se perdieron no es justificable por mero azar y no se tuvo en cuenta durante el análisis, los resultados estarán sesgados. Las técnicas de imputación múltiple de datos faltantes mitigan, bajo determinadas asunciones, la falta de potencia estadística debida a la pérdida de información, y además, corrigen los sesgos presentes en los análisis que

emplean exclusivamente a los participantes con todos sus datos completos (*complete-case analysis*) (7–13).

Las asunciones sobre el patrón de pérdida de datos (*missingness pattern*) es un aspecto crítico de cara a realizar una imputación múltiple de datos faltantes. Little y Rubin (14) enmarcaron tres situaciones posibles:

- a) Pérdida de datos completamente aleatoria (*missing completely at random* –MCAR–). Bajo esta asunción, la probabilidad de pérdida de datos no depende de los datos que se están observando ni tampoco de aquellos que no se observan.
- b) Pérdida de datos aleatoria (*missing at random* –MAR–). La probabilidad de pérdida de datos no depende de aquellos datos que no se observan, pero sí de aquellos que se están observando (y por tanto, recogiendo durante la investigación).
- c) Pérdida de datos no aleatoria (*missing not at random* –MNAR–). La probabilidad de pérdida de datos depende de datos que no se están observando (ni por tanto, recopilando).

Por ejemplo, en un estudio que pretenda evaluar la presión arterial, si algunos pacientes no acuden a la revisión, o un día el esfigmomanómetro no funciona y la medición no es válida, o el médico está de baja ese día, se podrá asumir un patrón de pérdida de datos completamente aleatoria (MCAR) en la variable “Presión arterial”. En cambio, como la hipertensión está correlacionada positivamente con la edad, es más probable que se ponga mayor atención en recoger datos de presión arterial de personas más mayores que de jóvenes y por tanto, habría un mayor número de datos faltantes entre los jóvenes. Bajo esta asunción, el patrón de pérdida de datos es el de pérdida de datos aleatoria (MAR), la probabilidad de perder los datos sobre presión arterial depende de la edad y ésta se recoge en la investigación. Por último, si la pérdida de datos de presión arterial depende de la edad, pero esta no se recogiera en la investigación, el patrón que debemos asumir es el de pérdida de datos no aleatoria (MNAR). En la práctica, no es posible distinguir entre MAR y MNAR solamente por los

datos observados, aunque la asunción MAR puede ser más plausible recogiendo un mayor número de variables explicativas e incluyéndolas en el análisis.

Asumir un patrón de pérdida de datos completamente aleatoria (MCAR), implica una pérdida de potencia en el análisis, pero los resultados no estarán sesgados si no se realiza una imputación múltiple de datos faltantes. En cambio, llevar a cabo imputación múltiple bajo la asunción MAR, corregirá esos sesgos si las variables que explican la pérdida de datos se tienen en cuenta durante la imputación. La solución frente a un patrón MNAR es compleja y no está resuelta a día de hoy. Por tanto, es fundamental tener en cuenta estos aspectos en la fase de diseño para poder asumir un patrón MCAR o MAR antes de realizar imputación múltiple.

Mediante imputación múltiple se reemplaza cada valor perdido por un conjunto de valores simulados con el fin de incorporar a la estimación la incertidumbre debida a la presencia de datos faltantes (15). La técnica de imputación múltiple de datos faltantes por ecuaciones encadenadas (*multiple imputation by chained equations-MICE-*) es una estrategia práctica y flexible en el manejo de los datos faltantes. Fue desarrollada y descrita en el ámbito de la medicina por van Buuren en 1999(16), y actualmente es una de las más utilizadas en investigación biomédica (10). Consiste en asignar valores a los datos faltantes múltiples veces, creando múltiples conjuntos de datos “completos”. Los valores perdidos son imputados a partir de valores observados para un individuo dado y las relaciones observadas en los datos de otros participantes, asumiendo que las variables que explican la pérdida de datos están incluidas en el modelo. Se incorporan componentes aleatorios en los valores estimados para reflejar su incertidumbre. Una vez se han obtenido los múltiples conjuntos de datos, se analizan individualmente de forma idéntica para obtener un conjunto de estimadores de los parámetros. Finalmente, los estimadores se combinan y se consiguen los estimadores globales, sus varianzas e intervalos de confianza (9,17).

La imputación múltiple por ecuaciones encadenadas, está basada en un conjunto de modelos de imputación: se ejecuta un modelo de regresión para cada variable que contenga datos faltantes, condicionando ese modelo a los demás datos. Los valores imputados se predicen a partir de esos modelos, añadiendo un error aleatorio. El

algoritmo es el siguiente: Inicialmente, todos los valores perdidos son sustituidos por valores extraídos aleatoriamente con re-emplazamiento del conjunto de valores observados para cada variable. A continuación, sobre la primera variable con el menor número de datos faltantes (llamada X_1 , por ejemplo), se aplica una regresión condicionando el resto de variables (X_2, X_3, \dots, X_k) como términos independientes, sólo sobre aquellas observaciones donde X_1 esté completa. Los valores faltantes de X_1 son reemplazados entonces por valores simulados (*simulated draws*) de la correspondiente distribución predictiva posterior de X_1 . Sobre la siguiente variable en orden ascendente con datos faltantes, X_2 , se aplica entonces otra regresión con todas las demás variables como términos independientes (X_1, X_3, \dots, X_k), restringiendo esta regresión a aquellas observaciones con todos los valores de X_2 observados y utilizando los valores imputados de X_1 . De nuevo, los valores perdidos de X_2 son reemplazados por valores simulados (*draws*) de la distribución predictiva posterior de X_2 . Este proceso se repite sucesivamente para todas las demás variables con datos faltantes. A esto se le conoce como “ciclo”. Para estabilizar los resultados, el procedimiento suele repetirse varios ciclos, normalmente 10 o 20 para producir un único archivo de datos imputado promediando los resultados individuales. El proceso completo se repite m veces para conseguir m conjuntos de archivos de datos con datos imputados (sin datos faltantes) en todos ellos. Una característica destacada de MICE reside en su habilidad para el manejo de diferentes tipos de variables (continuas, binarias, categóricas ordenadas o desordenadas), ya que cada variable es imputada a partir de su propio modelo de imputación (9).

Los modelos pronósticos se basan en la combinación de características de las observaciones para predecir resultados. A veces, la aplicación de estos modelos puede resultar en modelos que se ajustan pobremente a la realidad, o que predicen inadecuadamente el resultado sobre nuevas observaciones. Además, el uso de estrategias de modelización (*forward*, *stepwise* o *backward*), tiende a proporcionar modelos sobre-optimistas (6). Por ello, es crítico evaluar el rendimiento de un modelo predictivo (*accuracy*). En este sentido y en el ámbito de los modelos logísticos o de Cox, deben valorarse dos aspectos, calibración y discriminación. La calibración es la propiedad del modelo para realizar predicciones que se ajusten a la realidad. La

discriminación, es la capacidad que tiene el modelo para distinguir entre aquellas observaciones en las que ocurre el resultado de interés de aquellas en las que no. Desgraciadamente, la presencia de estos aspectos en la literatura científica es escasa (4,18–21).

Un artículo publicado recientemente por Hemingway et al.,(2) cita que “la investigación en pronóstico debe ser reconocida como un campo de investigación relevante en la investigación traslacional, e intrínseca a la práctica de la medicina clínica y el desarrollo de políticas sanitarias”.

Trasladando la aplicación de los modelos pronósticos al campo biomédico en que las observaciones se traducen en pacientes, debe evaluarse si el modelo se comporta igual en otros pacientes, diferentes de aquellos sobre los que se ha desarrollado. Esto es lo que se conoce como validación externa (*generalizability*), y está formada por dos aspectos a valorar, la capacidad de realizar predicciones válidas sobre una nueva muestra perteneciente a la misma población (*reproducibility*) y la capacidad de generar predicciones válidas sobre una población diferente, pero plausiblemente relacionada (*transportability*) (22). A menudo, ocurre que el fracaso de un modelo en su validación externa se podría haber predicho a partir de una rigurosa validación interna (no sesgada). Es decir, es probable que en muchos modelos cuya validación externa fracasara, se podría haber encontrado que también fallaba en series de sujetos de la muestra original, debido a un sobre-ajuste. Los tres métodos principales para obtener una evaluación de la validez interna prácticamente sin sesgos, son la división de la muestra de datos o *data-splitting*; la validación cruzada o *cross-validation* y las técnicas de re-muestreo o *bootstrapping*.

En el *data-splitting*, la muestra se divide en dos, por ejemplo dos tercios de la muestra global en el llamado grupo de trabajo o entrenamiento y un tercio en el grupo de validación. Aunque habitualmente la división se ha realizado de forma aleatoria, la recomendación actual, es hacerlo bajo algún criterio, por ejemplo, centro de estudio o tiempo para así reducir la similitud entre ambos grupos (3). Sobre uno de los grupos (el de trabajo) se desarrolla el modelo y sobre el otro (grupo de validación) se valida. La técnica de validación cruzada o *cross-validation* consiste en realizar sucesivas

repeticiones de *data-splitting*. Es decir, se repite el proceso de dividir la muestra en grupo de trabajo y validación, desarrollar el modelo y validarlo n veces, promediando los resultados sobre las n repeticiones, dejando fuera en cada repetición a un número determinado de observaciones. La técnica de *bootstrapping* consiste en tomar un gran número de muestras con reemplazamiento de la muestra original (de forma que se mantenga el tamaño original de la muestra) e ir ajustando el modelo en cada una de ellas. De esta manera, se obtienen estimadores prácticamente insesgados de la exactitud predictiva del modelo con una varianza relativamente pequeña.

La técnica de *bootstrapping* es considerada la más estable de las tres, puesto que se ha demostrado que las técnicas de *data-splitting* y *cross-validation* pueden ofrecer estimadores muy diferentes cuando el proceso completo de validación se repite varias veces (6,23).

Hasta hace unos años, estas técnicas avanzadas de análisis, la imputación múltiple de datos faltantes y la validación de los modelos por técnicas de *bootstrapping*, no se encontraban disponibles para la mayoría de los investigadores. Se necesitaban ordenadores con gran capacidad de procesamiento para llevar a cabo estos algoritmos. Actualmente, se encuentran disponibles en los paquetes estadísticos más utilizados en investigación biomédica, como *Stata* y *R*, pero se sigue requiriendo conocimiento experto para llevarlos a cabo. Con esta Tesis, se pretende abordar y avanzar en el uso de estas técnicas de análisis utilizando como ejemplo el registro de datos de la Red Española de Investigadores en Dolencias de la Espalda (REIDE).

La lumbalgia inespecífica se define como el dolor localizado entre el límite inferior de las costillas y el límite inferior de las nalgas, que no está causado por fracturas, traumatismos directos o enfermedades sistémicas, en el que no existe una compresión radicular demostrada y subsidiaria de tratamiento quirúrgico (24).

Con relación al tratamiento de la lumbalgia inespecífica, la Guía de Práctica Clínica de referencia para su manejo pertenece al Programa Europeo COST B13 (25,26), cuya versión española se publicó en el año 2005 (27). Está adoptada desde entonces por la Sociedad Española de Medicina General, Sociedad Española de Radiología, Red Española de Investigadores en Dolencias de la Espalda, Red Temática de Medicina

Basada en la Evidencia y Centro Cochrane Iberoamericano, entre otras; y recomendada por la Organización Médica Colegial. En esta Guía, entre los tratamientos que recomiendan para el manejo de la lumbalgia, se encuentra evitar el reposo en cama, mantener el mayor grado posible de actividad física, utilizar como fármacos de primera línea analgésicos, anti-inflamatorios y miorrelajantes y la neurorreflejo terapia (NRT).

En los países industrializados, se estima que, entre el 70% y el 84% de la población adulta, en algún momento de su vida padece dolor lumbar (28,29). En su manejo, es necesario un abordaje desde varios ámbitos. Por un lado, supone un problema de salud a gran escala, dada su alta prevalencia en las sociedades industrializadas. También juega un importante papel a nivel social, debido a la pérdida de calidad de vida y aumento en los niveles de discapacidad para desempeñar actividades de la vida cotidiana. Y por supuesto, una importante carga económica, consecuencia de los costes derivados de bajas laborales, tratamientos farmacológicos y rehabilitadores (30). Según un informe realizado por la Sociedad Española del Dolor en el año 2010, los costes del tratamiento del dolor lumbar en España equivalen al 2% del Producto Interior Bruto (PIB), ya sean directos (recursos humanos, coste de materiales y técnicas, camas hospitalarias, gasto farmacéutico, etc.) o indirectos (bajas laborales, pensiones, indemnizaciones, etc.).

Además, hay evidencias de que la carga socioeconómica del dolor lumbar crónico está incrementándose en los últimos años, a pesar de los avances tecnológicos en diagnóstico y la introducción de numerosas intervenciones para tratarlo (2,31).

Parte de la dificultad en su manejo reside en que el dolor lumbar es un síntoma y no una enfermedad y que en alrededor del 90% de los casos, no es posible identificar la causa exacta del dolor (26). Esto implica una barrera adicional de cara a entender su mecanismo y, derivado de ello, una gran dificultad sobre qué intervención específica ayudaría a un paciente en concreto (32).

Por todo esto, es importante desarrollar modelos que puedan predecir qué variables se asocian a la mejoría del dolor y discapacidad en aquellos pacientes que padecen lumbalgia inespecífica, al mismo tiempo que trasladar a la práctica clínica los

resultados de estos modelos, de forma que puedan ser fácilmente interpretados y aplicados por cualquier profesional de la clínica y por los propios pacientes.

INTRODUCTION

Prognosis means a foretaste, a prediction or risk estimation for future situations. In medicine, the concept of prognosis is associated with the risk that an individual achieves a health status or, an outcome over a specific period of time based on his clinical also not clinical profiles.

Although in this context, risk and prognostic factors could be interpreted as equivalents; there are some conceptual differences between them. A risk factor can be observed in a healthy population and its presence being associated with a higher chance of developing a disease. Therefore, a risk factor represents the probability of occurrence of an epidemiologic fact. By contrast, a prognostic factor, needs a departure from a disease situation, and predicts its evolution once the disease has appeared (1,2).

Since there is a great variability between patients, etiologies, disease onsets and prescribed treatments, only one prognostic factor will hardly provide an appropriate estimation of prognosis. Hence, prognosis studies need to use several predictors in order to estimate the disease evolution in a patient. To estimate which relevant predictors are associated with an outcome, likelihoods that combine different predictors, and tools that facilitate the estimation of these probabilities, a multivariable approximation is required, both among the design and analysis phases. These tools are the prognostic models (3,4).

The starting-point for a prognosis study must be the collection of data over the time axis. It is feasible to design cohorts studies, case-control or clinical trials, according to the objectives established by researchers. If research is focused on the unbiased observation between exposure and outcome besides estimating incidences, the best design is a prospective cohort. This allows a better approximation to the “natural history” of disease, having a good control over selection of subjects and measurements, also moving along a scenario as similar as possible to the clinical practice.

Interestingly, despite similarities in the design and analyses phases between prognostic models and aetiological research, predicting an outcome does not imply explaining its causality. Prognostic studies do not try to explain causal inference between a factor

and a result, but use several variables in order to accurately predict the risk of future events. Association does not imply causation. Estimators usually employed in a prognosis research (odds ratio, relative risk, hazard ratio), are not necessarily causal. They are only used to obtain a probability of the outcome for an individual. Aetiological studies focus on the risk of an outcome in presence of a causal factor relative to the risk in its absence (3,5).

To perform a prognosis research, the first step is developing the prognostic model, then validating its performance and finally, assessing its clinical impact. Multiple regression modeling allows the estimation from a set of variables the probability of a patient improving his disease along a period of time (6). According to dependent variable characteristics, the model can be linear or logistic, and if time to event is relevant, the chosen model should be a proportional hazard Cox model.

If the objective is to predict which variables are associated with a binomial variable, the best model is logistic regression. It fits into Generalized Linear Model (GLM), and uses the *logit* function as link function. Probability that describes the outcome is modeled as a function of the independent variables by means of a logistic function.

Any experimental or observational research study is exposed to missing data. This implies a lack of power in the analysis of results, because the traditional trend has been to ignore those observations with missing data. Furthermore, if the reason for missing data was beyond chance, and it was not taken into account during the analysis, results probably will be biased. Multiple imputation techniques mitigate, under determined assumptions, the lack of statistical power, and also correct bias present in completed cases analyses (7–13).

A critical aspect in multiple imputation of missing data is the requirement of some assumptions about missing data patterns. Little and Rubin(14) define three possible situations:

- a) Missing completely at random (MCAR). Under this assumption, the probability of missing data does not depend on observed neither unobserved data.

- b) Missing at random (MAR). The probability of missing data does not depend on unobserved data, but they do on observed (therefore, they are recorded along the research).
- c) Missing not at random (MNAR). The probability of missing data depends on unobserved data (therefore, they have not been recorded along the study).

For instance, a study of blood pressure, if some patients do not attend their visit, or the sphygmomanometer does not work properly so the measure is not valid, or the doctor is on sick leave that day, it could be assumed a missing completely at random (MCAR) pattern at the “blood pressure” variable. In another way, as having high blood pressure is correlated with age, is likely to pay more attention recording data of blood pressure in older people than younger, so there will be more missing data among young people. Under this assumption, the missing data pattern fits in a missing at random (MAR), i.e. the probability of missing data about blood pressure depends on age and it is recorded in the study. At last, if missing data of blood pressure depends on age, however, it is not recorded in the study, the missing data pattern assumed would be a missing not at random (MNAR) pattern. In practice, it is not possible to distinguish between MAR or MNAR patterns only looking at the observed data, although MAR assumption can be more plausible collecting a great deal of explicative variables and including them into the analysis.

Assuming a missing completely at random (MCAR) pattern implies a lack of power in the analysis. However, results will not be biased if a multiple imputation of missing data is not performed. On the other hand, carrying out a multiple imputation under MAR assumption, will correct bias if variables that explain the probability of missing are taken into account for the imputation process. The solution in case of MNAR pattern is complex and it is not solved currently. So therefore, it would be desirable to take into account these aspects during the design stage in order to be able of assume a MCAR or MAR pattern before performing a multiple imputation analysis.

By means of multiple imputation, each missing value is replaced by a set of plausible values to incorporate the uncertainty caused by missing data to the estimation (15). Multiple imputation by chained equations (MICE) is a practical and flexible strategy in management of missing data. It was developed and also described in the medicine field by van Buuren in 1999(16), and nowadays is one of the most used techniques in biomedical research (10). It consists of assigning values to missing data multiple times, creating multiple “complete” datasets. Missing values are imputed from observed values from a given individual and the observed relationships in data from other participants, assuming that variables which explain the missing data are included into the model. Random components are incorporated into these estimated values to reflect their uncertainty. Multiple datasets are created and then analyzed individually but identically to obtain a set of parameter estimates. Finally, the estimates are combined to obtain the overall estimates, variances and confidence intervals (9,17).

Multiple imputation by chained equations is based on a set of imputation models: a regression model is performed for each variable with missing values, conditioning this model to the rest of data. Imputed values are predicted from these models, adding a term of random error. The algorithm is as follows: initially, all missing values are filled in by simple random sampling with replacement from the observed values. Then, the first variable with missing values (for instance, X_1) is regressed on all other variables (X_2, X_3, \dots, X_k), restricted to individuals with the observed X_1 . Missing values in X_1 are replaced by simulated draws from the corresponding posterior predictive distribution of X_1 . Then, the next variable with missing values (X_2 say), is regressed on all other variables (X_1, X_3, \dots, X_k), restricted to individuals with the observed X_2 , and using the imputed values of X_1 . Again, missing values in X_2 are replaced by draws from the posterior predictive distribution of X_2 . The process is repeated for all other variables with missing values in turn: this is called a cycle. In order to stabilize the results, the procedure is usually repeated for several cycles, normally 10 or 20 times, to produce a single imputed dataset and the whole procedure is repeated m times to give m imputed datasets. An important feature of MICE is its ability to handle different variable types (continuous, binary, unordered categorical and ordered categorical) because each variable is imputed using its own imputation model (9).

Prognostic models are based on the combination of observed characteristics in order to predict results. Sometimes, the application of these models can result in models that poorly reflect reality, or predict the outcome inadequately along new observations. Moreover, the use of modeling strategies (*forward, stepwise or backward*), tends to give optimistic models (6). For that reason, is critical to assess the predictive model *accuracy*. In the field of logistic and Cox regression models, two aspects must be assessed: calibration and discrimination. *Calibration* refers to the ability of the model to make predictions which fit to reality. *Discrimination* refers to the ability of the model to distinguish between those observations which the event of interest occurs and those who do not. Unfortunately, the presence of these aspects in the scientific literature is scarce (4,18–21).

In a recent publication, Hemingway et al.,(2) recommends that “prognosis research should be recognized as a field of inquiry important in translational research and intrinsic to the practice of clinical medicine and development of healthcare policy”.

Transferring the applicability of prognostic models into the biomedical field, the observations would divert into patients, and it should be assessed if the model provides accurate predictions in a different sample of patients. This is called *generalizability* and consists of two aspects, the ability to produce accurate predictions among patients not included in the development of the model but from the same population, that is called *reproducibility*; and the ability to produce accurate predictions in a sample drawn from a different but plausibly related population, *transportability* (22). Sometimes, the failure of model generalizability could have been predicted from a rigorous accuracy evaluation (unbiased). It is likely that many clinical models which failed to validate would have been found to fail on another series of subjects from the original source, because overfitting is such a common problem. There are three principal methods in order to obtain a nearly unbiased internal assessment of accuracy, the *data-splitting*; the *cross-validation* and *bootstrapping*.

In *data-splitting*, the sample is split in two parts, for instance 2/3 of the sample in the work or training set and a 1/3 at the test set. Despite the splitting has been traditionally at random, the current recommendation is performing it with specified

criteria, as by centre or time, as it reduces the similarity of the two sets of patients (33). Over the training set the model is developed and over the test set is validated. *Cross-validation* is repeated *data-splitting*. In other words, the process of splitting the sample into training and test sets is repeated, as well as developing the model, validating it n times and averaging the results into the n replays, leaving out an specific number of observations in each replay. *Bootstrapping* involves taking a large number of samples with replacement from the original sample (maintaining the original sample size). Thus it is provided nearly unbiased estimates of predictive accuracy that are of relatively low variance.

Bootstrapping is considered the most stable method, because *data-splitting* and *cross-validation* have shown a high variation of accuracy estimates when the entire validation process is repeated (6,23).

Until a few years ago, these methods of advanced analysis, as the multiple imputation of missing data and model validation by *bootstrapping*, were not available to most of researchers. A great computer processing was needed to perform the algorithms. Currently, they are available in most statistical packages used in biomedical research, as *Stata* and *R*, but it still requires expert knowledge to accomplish. Throughout this thesis, is intended to address and advance the use of these analysis techniques using the data record of the Spanish Back Pain Researchers Network (REIDE).

Non-specific low back pain (LBP) is defined as the pain in the area between the bottom of the rib cage and the buttock creases, which it is not caused by fractures, direct trauma or systemic disease (24).

Regarding the treatment for nonspecific LBP, the main Clinical Practice Guideline for the management to common LBP belongs to the European Program COST B13(25,26), with the corresponding Spanish version published in 2005 (27). It has been adopted by the Sociedad Española de Medicina General, Sociedad Española de Radiología, Red Española de Investigadores en Dolencias de la Espalda, Red Temática de Medicina Basada en la Evidencia y Centro Cochrane Iberoamericano, among others; and it is recommended by the Organización Médica Colegial. In this guide, among recommended treatments for management of LBP, include avoiding bed rest, advise

patients to stay active and continue normal daily activities including work if possible, prescribe medication as analgesics, NSAIDs and muscle relaxants and the neuroreflexotherapy (NRT).

In developed countries is estimated that, between the 70% and 84% of adult population, suffer at any time in their life a LBP (28,29). In its management, is necessary an approach from several areas. On the one hand, it represents a major health challenge, since its high prevalence among industrialized countries. On the other hand, it implies a social challenge, because it causes a loss of quality of life and an increase in disability levels. As a consequence, emerges into a great economic burden, because of the costs of sick leaves, pharmacological drugs and rehabilitation treatments (30). According a Sociedad Española del Dolor report published in 2010, costs for treatment of LBP in Spain are equivalents to 2% Gross Domestic Product (GDP), whether direct costs (human resources, techniques and material costs, hospital stays, pharmaceutical costs, etc) as well as indirect (sick leaves, pensions, compensations, etc.).

Moreover, there is evidence that economic burden for LBP is increasing over the last years, in spite of technological advances in diagnosis and the introduction of several interventions for its treatment (2,31).

Part of the difficulty in handling is that LBP is considered more as a symptom than a disease and about 90% of cases; it is not possible to identify the exact cause of pain (26). This implies an additional barrier in order to understand its mechanism and, derived from it, a great difficulty about what specific intervention will helps a particular patient (32).

For all these reasons, it is quite important to develop predicting models in order to know which variables are associated with the improvement of pain and disability in patients with LBP, at the same time transferring results from these models into the clinical practice, the way it can be easily interpreted and applied by anyone from clinical practice or patients themselves.

HIPÓTESIS

Si bien en esta tesis no hay una hipótesis formal a contrastar, subyacen varias hipótesis conceptuales orientadas a la aplicación y optimización de metodologías ya conocidas, a la práctica clínica:

- El empleo de métodos estadísticos avanzados (imputación múltiple de datos faltantes, modelos de regresión múltiple y técnicas de *bootstrapping*), permitirá conocer de forma válida qué variables se asocian a la evolución de los pacientes con lumbalgia y cervicalgia inespecífica tres meses después de una primera visita al médico.
- El desarrollo de herramientas que faciliten la interpretación de los resultados del análisis estadístico, permitirá al profesional un mejor manejo del dolor lumbar, cervical y discapacidad de sus pacientes.

OBJETIVOS

- Emplear técnicas de imputación múltiple de datos faltantes para corregir la falta de potencia y el posible sesgo derivado de recoger datos de manera incompleta.
- Desarrollar modelos pronósticos de mejoría de dolor lumbar, local e irradiado, en pacientes con lumbalgia inespecífica en el momento de la primera consulta.
- Desarrollar modelos pronósticos de mejoría de la discapacidad asociada a dolor lumbar en pacientes con lumbalgia inespecífica en el momento de la primera consulta.
- Desarrollar modelos pronósticos de mejoría de dolor lumbar y cervical, y dolor irradiado a la pierna o los brazos, en pacientes con lumbalgia y cervicalgia inespecífica tras una intervención no quirúrgica.
- Desarrollar modelos pronósticos de mejoría de la discapacidad asociada a dolor lumbar y cervical en pacientes con lumbalgia y cervicalgia inespecífica tras una intervención no quirúrgica.
- Emplear métodos de calibración y discriminación que evalúen el rendimiento de los modelos pronósticos.
- Validar los modelos pronósticos mediante técnicas avanzadas de re-muestreo (*bootstrapping*).
- Desarrollar herramientas, como nomogramas y una aplicación web, para facilitar la interpretación y aplicación de los resultados de los modelos a la práctica clínica.

ARTÍCULO 1:

**Predicting the evolution of low back pain patients in
routine clinical practice: results from a registry within
the Spanish National Health Service**

Justificación metodológica

En este primer artículo se plantean tres modelos predictivos para la mejoría del dolor lumbar, dolor irradiado a la pierna (ciática) y discapacidad asociada al dolor lumbar. Los datos de los participantes se recogieron gracias a un registro prospectivo de pacientes en centros especializados y de atención primaria, pertenecientes a la REIDE. En el registro, se incluyeron los datos de aquellos pacientes que dieron su consentimiento. Además, para participar en el estudio, se solicitó y obtuvo el consentimiento informado de los pacientes y la autorización de los correspondientes Comités de Ética de Investigación Clínica.

Hay varios aspectos metodológicos en este estudio que se han desarrollado con especial interés y cabe destacar en esta introducción. En primer lugar, la definición de cada una de las tres variables resultado u *outcomes* evaluadas en este estudio, que conllevan toda una metodología detrás denominada “Cambios mínimos clínicamente relevantes”.

Las bases que sientan precedentes para que esta metodología se desarrolle son, por un lado, que la significación estadística no es equivalente a relevancia clínica (34), y por otro, que el uso de instrumentos que miden la calidad de vida relacionada con la salud de los pacientes (CVRS) (*Health-Related Quality of Life* -HRQL-) es cada vez más común. La CVRS es un constructo multidimensional que se refiere a la medida en que unas expectativas se ven afectadas en cuanto a bienestar físico, emocional y social, por una condición médica o su tratamiento (35).

A diferencia de las medidas fisiológicas, como la presión arterial o el nivel de colesterol en sangre, el sentido del cambio en una unidad de una escala de CVRS es poco intuitivo, ya que las unidades no resultan tan comunes como los mm Hg o los mg/dl. Además, una diferencia de tan sólo una unidad, puede suponer un cambio diferente según se mida en diferentes grupos de pacientes con distintas enfermedades, niveles de intensidad, estado socioeconómico o nacionalidad. La diferencia mínima clínicamente importante se define como “la diferencia más pequeña en la escala de

interés que los pacientes perciben como beneficiosa y que supone, en ausencia de efectos secundarios perjudiciales y excesivos costes, un cambio en el manejo del paciente” (36). Determinar esta diferencia mínima clínicamente relevante es fundamental, ya que pequeñas diferencias numéricas, pueden implicar resultados estadísticamente significativos si el tamaño muestral es suficientemente grande y como ya se ha comentado anteriormente, significación estadística no implica relevancia clínica. El caso de las escalas de CVRS resulta de especial interés, dada la arbitrariedad de sus unidades.

La medición del dolor lumbar inespecífico y la capacidad funcional se encuentran incluidas en el grupo de medidas para evaluar la CVRS. Hasta el año 2007, no se había publicado nada al respecto en el ámbito de las dolencias de espalda en España. Por ello, la doctoranda junto a un equipo de investigación perteneciente a la REIDE, publicaron un estudio en el que se expresaba la magnitud de la diferencia mínima clínicamente relevante para la escala visual analógica (VAS) de dolor lumbar y discapacidad asociada a lumbalgia, medida por el cuestionario Roland Morris, y que se ha utilizado en este artículo para definir las variables resultado (37).*

Otro aspecto metodológico relevante de este artículo es el empleo de técnicas de imputación múltiple de datos faltantes por el método de las ecuaciones encadenadas (MICE), asumiendo un patrón de pérdida de datos al azar (MAR). Se ha realizado esta asunción partiendo del conocimiento clínico sobre la lumbalgia y los pacientes que la padecen, para poder introducir en el proceso de imputación múltiple aquellas variables que puedan explicar la probabilidad de la pérdida de datos. En el proceso de imputación de datos faltantes, se debe crear un determinado número de archivos de datos con los datos faltantes ya imputados. En el momento en que se llevó a cabo este estudio, no había consenso ni normas claras sobre cuántos archivos de datos debían crearse, y la tendencia era la de aplicar una regla empírica de entre 5 y 10 conjuntos de datos imputados (9,13,38).

*En el Apéndice I se adjunta la correspondiente publicación.

El desarrollo posterior de tres modelos múltiples de regresión logística fue con el objetivo de conocer qué variables se asocian a una mejoría clínicamente relevante en cuanto a dolor local, dolor irradiado y discapacidad asociada al dolor lumbar. Aunque algunas recomendaciones sugieren no realizar estrategias de modelización (6,23), en este primer artículo, y siguiendo el principio de parsimonia (39), se llevó a cabo una estrategia hacia atrás o *backward*, no automática. De las tres estrategias de modelización más utilizadas, hacia atrás o *backward*, hacia adelante o *forward* y paso a paso o *stepwise*, la más recomendada es la estrategia hacia atrás (20,33).

Aplicar los coeficientes de un modelo pronóstico para realizar una predicción, es laborioso. Además, el valor del coeficiente no es lo que realmente interesa cuando se habla de un modelo pronóstico, sino el valor del coeficiente multiplicado por el valor de la variable correspondiente. Frank Harrell señaló en el año 2000(23), que una buena manera de hacer interpretable un modelo multivariable era construir un nomograma que mostrara la asociación entre las variables predictoras y la respuesta. Un nomograma es un instrumento gráfico de cálculo, un diagrama bidimensional que permite el cómputo gráfico y aproximado de una función con cualquier número de variables. En su concepción más general, el nomograma representa simultáneamente el conjunto de las ecuaciones que definen un determinado problema y el rango total de sus soluciones. Citando textualmente a Harrell, "un nomograma no sólo arroja luz sobre cómo el efecto de un factor de predicción de la probabilidad de la respuesta depende de los niveles de otros factores, sino que además, permite estimar rápidamente la probabilidad de respuesta de los sujetos individuales". Es decir, que gracias al nomograma se obtiene directamente la probabilidad estimada de un evento a partir de los valores de sus variables predictoras. Esto es especialmente útil para usuarios que no dispongan de habilidades o herramientas de cálculo accesibles. De esta manera, los nomogramas se han convertido en herramientas cada vez más populares en la interpretación de los modelos predictivos.

El siguiente paso tras llevar a cabo un modelo de regresión, debe ser la valoración de su rendimiento, su calibración, la especificación del modelo y discriminación. Para la evaluación de la calibración del modelo, es ampliamente utilizada la prueba de bondad de ajuste de Hosmer-Lemeshow(40), que consiste en comparar las frecuencias de

eventos observados y predichos por el modelo, agrupando la muestra en deciles. Los modelos cuyas frecuencias observadas y esperadas son similares en todos sus deciles, serán modelos bien calibrados. Una de las estrategias para evaluar la especificación de un modelo, es la técnica de *bootstrapping*, basada en la obtención mediante técnicas de re-muestreo automático de un gran número de muestras-*bootstrap*, en las que, sobre cada una de ellas, se replicará la estrategia de modelización, para finalmente cuantificar la frecuencia de inclusión de cada variable en el modelo final (41).

Uno de los métodos más utilizados en la valoración de la discriminación de un modelo, es el área bajo la curva ROC(42). Se trata de un gráfico con la representación de la sensibilidad frente a (1 – especificidad) para un sistema clasificador binario según se varía el umbral de discriminación. En el caso de los modelos pronósticos, se construye para la probabilidad predicha por el modelo, y representa la probabilidad de que, dados dos individuos, uno que presentará el evento de interés y otro que no, el modelo asigne una mayor probabilidad de presentarlo al primero de ellos (33). El máximo valor posible del área bajo la curva ROC es igual a 1, representando que el modelo discrimina perfectamente, y 0,5 es el valor que indica que la discriminación del modelo es completamente aleatoria. Así, una mayor área bajo la curva ROC, supondrá una mejor discriminación del modelo.

Con este primer artículo, los objetivos que se persiguen son poner en marcha las técnicas de imputación múltiple de datos faltantes mediante el método de las ecuaciones encadenadas y desarrollar modelos válidos que predigan la probabilidad de mejorar de modo clínicamente relevante la intensidad del dolor lumbar, dolor irradiado a la pierna (ciática) y discapacidad asociada a dolor lumbar.

Participantes y métodos

En este artículo, se han utilizado datos de pacientes que acuden por lumbalgia inespecífica a su médico de cabecera o especialista. La REIDE invitó a participar en este estudio a 17 centros en siete provincias españolas. Trece de los centros pertenecen al Sistema Nacional de Salud (SNS) y cuatro a fundaciones sin ánimo de lucro adheridas al SNS. Los centros incluyen Centros de Salud de Atención Primaria y Centros de Especialidades en reumatología, rehabilitación, neurorreflejo terapia y cirugía ortopédica.

Los criterios de inclusión de los participantes fueron: acudir a uno de estos centros por lumbalgia, con o sin dolor irradiado a la pierna, que no esté causada por traumatismo o enfermedades sistémicas, no cumplir criterios para cirugía urgente y ser capaces de leer correctamente en castellano. Aquellos pacientes que habían sido operados de cirugía de la columna sin éxito y los que, a pesar de tener “señales de alerta”, se hubieran descartado enfermedades sistémicas, también se incluyeron.

Los criterios de exclusión fueron: pacientes que habían sido tratados o no por deterioro del sistema nervioso central y negativa a firmar el consentimiento informado.

El diseño del estudio corresponde a un estudio de cohortes prospectivo, con mediciones en la primera visita (visita basal) y a los tres meses. La elección de los tres meses como visita final se justifica por varias razones. En primer lugar, porque es el plazo recomendado en la valoración de un resultado para un episodio de lumbalgia, ya que asegura que todos los pacientes que comenzaron el estudio y pudieran seguir con dolor al cabo de este periodo, se haya cronificado (43). Para aquellos pacientes que ya entraron en el estudio padeciendo dolor crónico, la literatura científica apunta que su pronóstico se ve afectado por cambios en el dolor y discapacidad a lo largo de este tiempo (44). Además, estudios previos demuestran que, en el contexto del SNS, la pérdida de datos durante los 3 primeros meses de un estudio de investigación es mínima, aumentando considerablemente a partir del sexto mes de seguimiento.

En la primera visita, se recogieron las variables citadas en la Tabla 1.

Tabla 1. Variables recogidas en la visita basal

Características demográficas de los pacientes
-Género -Edad (años) -Situación laboral: Trabajando Baja laboral Estudiante Ama de casa Desempleado Retirado Otros
Variables relacionadas con el dolor lumbar
-Duración del episodio actual de lumbalgia (días) -Intensidad del dolor lumbar (escala VAS: 0-10) -Intensidad dolor irradiado a la pierna (escala VAS: 0-10) -Grado de discapacidad asociado al dolor lumbar (cuestionario Roland Morris: 0-24)
Características clínicas de los pacientes y antecedentes
-Cirugía lumbar previa (sí/no) -Síndrome de cirugía fallida (sí/no) -Procedimientos diagnósticos prescritos para este episodio: Rayos-X TAC Resonancia Magnética Electromiografía Otros (analítica sangre, escintigrafía) -Hallazgos radiológicos por este u otros episodios previos de lumbalgia: Deshidratación o degeneración discal, o artrosis Degeneración o artrosis de las articulaciones facetarias (o unciformes) Escoliosis Disimetría miembros inferiores Espondilólisis Espondilolistesis Estenosis espinal Fisura discal Protrusión discal (o hernia no extruida) Hernia discal (extruida) Lumbarización de S1 Sacralización de L5 Otros hallazgos radiológicos Ningún hallazgo radiológico
Tratamientos prescritos por este episodio
-Farmacológicos: Analgésicos Antiinflamatorios no esteroideos (AINEs) Corticoides Relajantes musculares

Opioides
Otros fármacos
-No farmacológicos:
Fisioterapia
Rehabilitación (ejercicio, técnicas de calor y/o frío, electroterapia, técnicas manuales como masaje o movilización)
Intervención NRT
Cirugía

Las tres variables principales se definieron como “Mejoría en el dolor lumbar”, “Mejoría en el dolor irradiado a la pierna” y “Mejoría en la discapacidad asociada a dolor lumbar”. En estas tres definiciones, se utilizó como punto de corte el correspondiente cambio mínimo clínicamente relevante. En estudios previos, se había fijado que este cambio, en términos generales, corresponde al 30% de la puntuación basal, con un valor mínimo de 1,5 puntos en la escala VAS y 2,5 puntos en la escala de Roland Morris (37,45). Con estas definiciones, en aquellos pacientes que tuvieran una puntuación basal, ya sea en intensidad de dolor o en discapacidad, inferior a estos puntos de corte, no es posible observar una mejoría clínicamente importante en ellos, pero sí un empeoramiento. Estos pacientes sólo se incluyen en el análisis si empeoraron. Es decir, en aquellos pacientes que, por ejemplo, no presentaran dolor irradiado al inicio del estudio, no puede observarse una mejoría clínicamente relevante y se excluyeron del análisis, salvo que aparezca dolor irradiado al final del estudio. Esto se aplica independientemente a cada uno de los modelos.

Dentro del manejo de los datos faltantes en este estudio, se llevó a cabo una imputación múltiple de datos faltantes por el método de las ecuaciones encadenadas, asumiendo un patrón en la pérdida de datos aleatorio (MAR). Esta asunción se basa en que en el registro se recogieron aquellas variables que explican el porqué de los datos faltantes, y por tanto, se incluyeron en el algoritmo de imputación junto a la variable dependiente (en cada caso). Se generaron 5 archivos de datos imputados para cada modelo (5 archivos de datos para el modelo de mejoría del dolor lumbar local, 5 para el modelo de mejoría del dolor irradiado a la pierna y 5 archivos para el modelo de mejoría de la discapacidad). Se aplicó una estrategia de modelización de forma independiente para cada archivo de datos, de modo que se obtuvieron 5 modelos finales para mejoría del dolor lumbar local, 5 modelos finales para mejoría del dolor

irradiado a la pierna y 5 modelos finales para la mejoría de la discapacidad. Para englobar los resultados de los 5 modelos finales en uno solo, el criterio fue incluir una variable en el modelo global, si permanecía al menos en dos de los 5 modelos finales correspondientes (46). Este criterio es diferente del aplicado en el segundo trabajo de esta tesis, porque en el momento en que se llevó a cabo este estudio, no se disponía de la herramienta apropiada. Se emplearon las técnicas de Rubin en la obtención del promedio de los coeficientes y sus errores estándar (12).[†]

Bajo criterios clínicos y estadísticos se seleccionaron un gran número de variables candidatas a introducir en el modelo máximo. La investigación en pronóstico posee una visión holística de todas las comorbilidades que siente una persona, mientras que otras áreas del conocimiento, como el diagnóstico implica un enfoque en un solo sistema de órganos o patología. El pronóstico de algunos tipos de cánceres, lesiones cerebrales o dolor lumbar están fuertemente influenciados por condiciones no relacionadas con el tumor, el cerebro o la columna, respectivamente (2). Por esta razón, el número de posibles candidatas a variables predictoras en la mejoría del dolor lumbar local e irradiado, y de la discapacidad, se encontraba en alrededor de la treintena.

El tamaño muestral no supuso un problema, puesto que el gran número de pacientes registrados en la REIDE, hicieron posible que el desarrollo de los modelos no se viera afectado por una posible pérdida de potencia o un riesgo de saturación del modelo.

Se incluyeron como variables independientes, además de las descritas en la Tabla 1 de esta misma sección, el ámbito de reclutamiento del paciente (atención primaria o especializada). Algunas variables se re-codificaron para evitar inestabilidad numérica y facilitar la interpretación de los modelos. Esto afectó a las variables:

- Duración del episodio actual, se clasifica en Agudo (menos de 14 días de duración), Sub-agudo (entre 14 y 89 días) y Crónico (igual o más de 90 días) (43,47).

[†] En el Apéndice II se encuentra una especificación detallada de las reglas de Rubin.

- Situación laboral, que se introdujo en los modelos de regresión dividida en tres categorías: Trabajador en activo, Trabajador de baja laboral por lumbalgia y Laboralmente no activo (agrupando las categorías de Estudiante, Ama de casa, Desempleado, Jubilado y Otros).
- Se fusionó la información que aportaban las variables Fisioterapia y Rehabilitación en una sola variable.

La estrategia de modelización fue hacia atrás y no automática, es decir, que de forma secuencial, se fueron eliminando aquellas variables que menos valor aportaban a la correspondiente “Mejoría”, fijando el punto de corte para mantener las variables en el modelo con un nivel de significación menor a 0,05.

Tras conseguir los tres modelos finales, se dibujaron tres nomogramas con sus respectivos coeficientes, utilizando el paquete estadístico R y sus librerías *rms*, *car*, *survival* y *Hmisc* (48).

En la evaluación de la especificación de cada modelo, se utilizó la técnica de *bootstrapping*. A partir de la generación de 1000 muestras-*bootstrap* de tamaño N (correspondientes a cada modelo), se replicaron las estrategias de modelización y se recogió la frecuencia de inclusión de cada variable en el modelo final (41). Este algoritmo se ejecutó de forma automática mediante un comando del software *Stata* (49), denominado *swboot*.

La valoración del rendimiento de los tres modelos, fue mediante evaluación de la discriminación y calibración. Para la discriminación se obtuvo el área bajo la curva ROC y la prueba de Hosmer-Lemeshow fue empleada en valorar la calibración del modelo.

Resultados

Los resultados presentados a continuación corresponden a la publicación del artículo *“Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service”*, publicado en la revista *The Spine Journal* en el año 2012.

Para facilitar la interpretación de los coeficientes de los tres modelos pronósticos, se incluyeron en la publicación sus correspondientes nomogramas. Atendiendo a las características de un paciente en concreto, se suman los puntos que correspondan resultando en una puntuación global, con la que se consigue una probabilidad de mejoría, tal y como se indica en las figuras 1, 2 y 3 del artículo publicado (páginas 1017-1018).

Con el fin de acercar los resultados de la investigación a la práctica clínica, viene demandándose desde hace unos años, la necesidad de fortalecer la investigación traslacional (50,51). Las acciones de traslación son todas aquellas que permiten la aplicación a la práctica clínica del conocimiento generado a través de la investigación.

Adquiere una especial relevancia por tanto, hacer efectivo el retorno social de la labor investigadora y, bajo esta deuda adquirida con la sociedad, se ha desarrollado una aplicación web, disponible en la siguiente dirección:

www.pronosticodolorlumbar.es

En esta página se encuentra una herramienta que es accesible tanto a pacientes como a profesionales, que estima la probabilidad de mejoría clínicamente relevante en el dolor lumbar, dolor irradiado a la pierna (ciática) y discapacidad asociada al dolor lumbar en los tres meses siguientes a una primera consulta.

El paciente o el médico, sólo tiene que introducir los valores de las variables solicitadas en la aplicación (Figura 1), seleccionar el botón de “Calcular” y la aplicación mostrará el resultado en cada caso (Figura 2).

La aplicación cuenta con una ayuda que sirve de tutorial para conocer a quién va dirigida esta herramienta, en qué está basado el cálculo de las probabilidades que se muestran (citando el presente trabajo de investigación) y sus limitaciones. También cuenta con una explicación a cada una de las preguntas que deben completarse para obtener el resultado.

A modo de ejemplo, para un paciente que lleva con dolor lumbar aproximadamente un mes, el día que acude al médico fija su intensidad de dolor lumbar en 7 en una escala visual analógica (VAS) que oscila entre 0 y 10 puntos, una intensidad de dolor irradiado igual a 3 (medido en la misma escala VAS que el dolor lumbar), un grado de discapacidad igual a 12 puntos en la escala Roland Morris (intervalo entre 0 y 24 puntos), sin cirugía lumbar previa, pero con hallazgo de degeneración discal en la última prueba radiológica realizada, y toma relajantes musculares para aliviar su dolor. Este paciente tendrá una probabilidad del 46.3% de mejorar su dolor lumbar de forma clínicamente relevante dentro de tres meses, un 28% en cuanto a mejorar el dolor irradiado a la pierna y un 34% para mejorar su discapacidad (Figura2). Con esta información, tanto el clínico como el paciente podrán conocer el pronóstico a los tres meses de la primera consulta.

En su primer año de funcionamiento, la página web ha recibido 13.460 visitas (periodo comprendido entre el 01 de enero de 2013 y el 01 de enero de 2014).

Figura 1. Visualización de la aplicación web a la hora de introducir información sobre las diferentes variables.

¿Hace cuántos días que comenzó el episodio doloroso que está padeciendo? ⓘ

☐ <14 días (agudo)
☒ 14-90 días (subagudo)
☐ >90 días (crónico)

¿De qué intensidad es el dolor lumbar que está padeciendo hoy? ⓘ

Ausencia de dolor Máximo dolor imaginable

7

¿De qué intensidad es el dolor irradiado a la pierna ("ciática") que está padeciendo hoy? ⓘ

Ausencia de dolor Máximo dolor imaginable

3

¿Cuál es el grado de incapacidad (o "restricción de las actividades cotidianas") que causa el dolor lumbar? ⓘ

12

¿Le han operado alguna vez de la columna lumbar? ⓘ

☐ Sí
☒ No

¿Le han diagnosticado "degeneración discal"? ⓘ

☒ Sí
☐ No

¿Le han prescrito un "electromiograma" por el episodio doloroso que está padeciendo? ⓘ

☐ Sí
☒ No

Para tratar el episodio doloroso que está padeciendo, ¿le han recetado relajantes musculares? ⓘ

☒ Sí
☐ No

Para tratar el episodio doloroso que está padeciendo, ¿le han realizado una intervención neuroreflejojoterápica? ⓘ

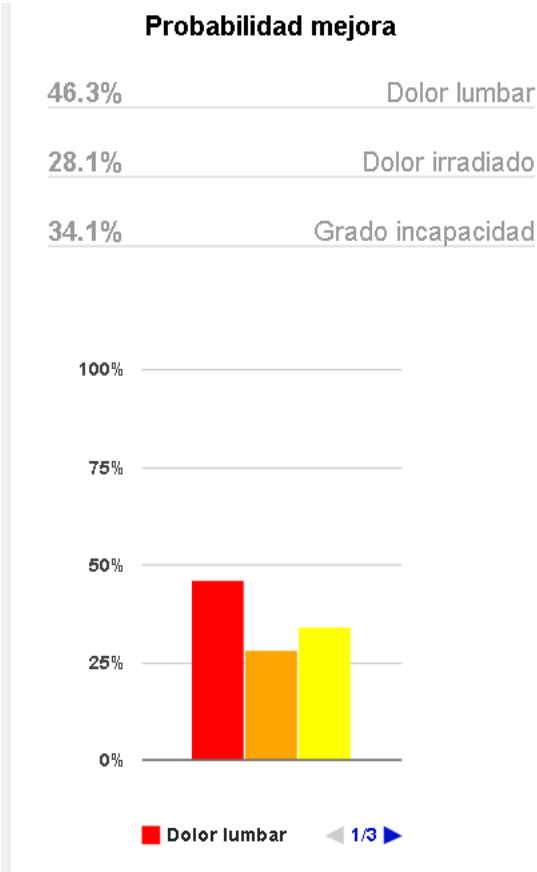
☐ Sí
☒ No

Para tratar el episodio doloroso que está padeciendo, ¿le han recetado opiáceos (o "mórficos")? ⓘ

☐ Sí
☒ No

Calcular

Figura 2.Visualización de la aplicación web mostrando la probabilidad de mejoría en los tres modelos para un caso concreto.



Clinical Study

Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service

Francisco M. Kovacs, MD, PhD^{a,b,*}, Jesús Seco, MD, PhD^{c,b}, Ana Royuela, MSc^{d,e,b}, Josep Corcoll Reixach, MD^f, Víctor Abaira, PhD^{e,d,b}, and the Spanish Back Pain Research Network

^aDepartamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain

^bSpanish Back Pain Research Network, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain

^cDepartamento de Enfermería y Fisioterapia, Instituto de Biomedicina, Universidad de León, Campus de Ponferrad, Avda. de Astorca s/n, 24401 Ponferrada, Spain

^dCIBER Epidemiología y Salud Pública (CIBERESP), C/ Melchor Fernández Almagro 3-5, 28029 Madrid, Spain

^eUnidad de Bioestadística Clínica, Hospital Ramón y Cajal, IRYCIS, C/ Colmenar Viejo, 9, 28031 Madrid, Spain

^fCentro de Salud de Tramuntana-Esporlas, C/ Coarter 24, 07019 Esporlas, Spain

Received 8 March 2011; revised 16 August 2012; accepted 9 October 2012

Abstract

BACKGROUND CONTEXT: The Spanish National Health Service (SNHS) is a tax-funded public organization that provides free health care to every resident in Spain.

PURPOSE: To develop models for predicting the evolution of low back pain (LBP) in routine clinical practice within SNHS.

FDA device/drug status: Not applicable.

Author disclosures: **FMK:** Nothing to disclose. **JS:** Nothing to disclose. **AR:** Nothing to disclose. **JCR:** Nothing to disclose. **VA:** Nothing to disclose.

Other members of the Spanish Back Pain Research Network who authored this study are Andrés Peña (Servicio de Rehabilitación, Hospital Ramón y Cajal, Madrid, Spain), Alfonso Muriel, MSc (CIBER Epidemiología y Salud Pública [CIBERESP], Spain; Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain), María Nieves Plana (CIBER Epidemiología y Salud Pública [CIBERESP], Spain; Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain), Antonio Pallicer (Centro de Salud Santa Eulalia, Ibiza, Spain), Mario Gestoso (Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain), Nicole Mufraggi (Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain), Carlos Isanta (Centro de Salud San José Norte, Zaragoza, Spain), Monserrat Núñez (Hospital Clínic, Barcelona, Spain), José Luis Peña Sagredo (Hospital Universitario Marqués de Valdecilla, Santander, Spain), Pilar Brieva (Hospital Universitario Marqués de Valdecilla, Santander, Spain), Luis Alvarez Galovich (Fundación Jiménez Díaz, Madrid, Spain), Salvador Fuster (Hospital Clínic, Barcelona, Spain), Javier Zamora (CIBER Epidemiología y Salud Pública [CIBERESP], Spain; Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain), Daniel Benítez Merelo (Centro de Salud de Montequinto, Sevilla, Spain), Francisco Martínez (Centro de Salud Bembibre, León, Spain), Elena Rodríguez (Unidad Básica Bahía Grand, Lluçmajor, Balears, Spain), Rosa Vecino Lafort (Centro de Salud San José Norte, Zaragoza, Spain), Dolores Vázquez (Mutua Asepeyo, Madrid, Spain), Mariano Ortega (Centro de Salud Lluçmajor, Balears, Spain), Eva Allué (Centro de Salud San José Norte, Zaragoza, Spain), Teresa López-Melus (Centro de Salud

San José Norte, Zaragoza, Spain), Patricia Gracia (Centro de Salud San José Norte, Zaragoza, Spain), Ma Dolores Vicente (Centro de Salud San José Norte, Zaragoza, Spain), Ma Ángeles Gay (Centro de Salud San José Norte, Zaragoza, Spain), Isabel Torres (Centro de Salud San José Norte, Zaragoza, Spain), Pilar Borraz (Centro de Salud de Andorra, Teruel, Spain), Emilio Suárez Sánchez (Centro de Salud de Montequinto, Sevilla, Spain), Ana Expósito (Centro de Salud San José Norte, Zaragoza, Spain), Ana Hernández-Moreno (Centro de Salud Santo Grial, Huesca, Spain), Gloria Sanz (Centro de Salud San José Norte, Zaragoza, Spain), Yolanda Sánchez (Centro de Salud San José Norte, Zaragoza, Spain), Pilar Melón Juncosa (Centro de Salud San José Norte, Zaragoza, Spain), Mireia Barceló Castelló (Centro de Salud San José Norte, Zaragoza, Spain), José Sánchez Blanco (Distrito Sanitario Sevilla Sur), Margarita Martín (Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain), Domingo Carretero (Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain), Modesto Ferrer (Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain).

This study was funded by the Kovacs Foundation, a Spanish not for profit institution with no links to the health industry. The funding institutions had no role in the design and conduction of the study; data collection; management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or the decision to submit the article for publication. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this article.

* Corresponding author. Departamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain. Tel.: (34) 971720809.

E-mail address: kovacs@kovacs.org (F.M. Kovacs)

STUDY DESIGN: Analysis of a prospective registry in routine clinical practice, in 17 centers across SNHS.

PATIENT SAMPLE: Patient sample includes 4,477 acute and chronic LBP patients treated in primary and hospital care.

OUTCOME MEASURES: Pain and disability, measured through validated instruments.

METHODS: Patients treated for LBP were assessed at baseline and 3 months later. Data gathered were the following: sex, age, employment status, duration of pain, severity of LBP, pain down to the leg (LP) and disability, history of lumbar surgery, diagnostic procedures undertaken, imaging findings, and treatments used throughout the study period. Three separate multivariate logistic regression models were developed for predicting a clinically relevant improvement in LBP, LP, and disability at 3 months.

RESULTS: In total, 4,261 patients (95.2%) attended follow-up. For all the models, calibration was reasonable and the area under the receiver operating characteristic curve was ≥ 0.640 . For LBP, LP, and disability, factors associated with a higher probability of improvement at 3 months were the following: not having undergone lumbar surgery, higher baseline scores for the corresponding variable, lower ones for the rest, and being treated with neuroreflexotherapy. Additional factors were the following: for LBP, shorter pain duration; for LP, not undergoing electromyography; and for disability, shorter pain duration, not being diagnosed with disc degeneration, and being treated with muscle relaxants and not opioids.

CONCLUSIONS: A prospective registry can be used for developing predictive models to quantify the odds that a given LBP patient will experience a clinically relevant improvement. This may empower patients for an informed shared decision making. © 2012 Elsevier Inc. All rights reserved.

Keywords:

Low back pain; Disability; Prognosis; Routine clinical practice; Clinical outcome

Introduction

Randomized controlled trials (RCTs) are the gold standard for assessing the efficacy and effectiveness of treatments. However, results in routine clinical practice do not necessarily match those obtained in experimental conditions. Therefore, although high-quality RCTs are indispensable for deciding whether a technology should be introduced into clinical practice, they may not be valid for anticipating actual results within routine practice.

Registries can be valuable to determine the effects of treatments that have been previously shown to be effective through RCTs, in routine clinical practice; quantify the influence of treatments on patients' prognosis once other prognostic factors have been accounted for; compare the results of treatments shown to be effective for the same subset of patients; identify factors that predict the result of each treatment, to refine its indication criteria; gather data on potential safety concerns, which had been previously undetected [1–3]; empower patients for informed shared decision making by developing predictive models that provide them with data on the probability of improvement associated with clinical decisions, factoring in their personal and clinical characteristics.

Low back pain (LBP) represents a major cause of public and health expenditure in most developed countries. In the United States, health expenditure for LBP is similar to expenditure on cancer and grew 65% between 1997 and 2005 [4]. Data do not suggest that this increase in expenditure has led to an improvement in results [4]. This may be one of the reasons why one of the top priorities

recommended by the Institute of Medicine, for investing the \$1.1 billion devoted to comparative effectiveness research by the 2009 American Recovery and Reinvestment Act, was to “establish a prospective registry to compare the effectiveness of treatment strategies for low back pain without neurological deficit or spinal deformity” [5].

To be useful, registries must include virtually all the patients; gather valid, reliable, and clinically relevant data; and ensure that losses to follow-up are kept to a minimum. Concerns have been expressed with regard to the feasibility of these requirements in routine clinical practice [6].

Therefore, the objectives of this study were to explore the feasibility of implementing a registry of patients treated for LBP in routine practice and use it for developing predictive models to quantify the likelihood that a given patient experiences a clinically relevant improvement in LBP, LP, and disability.

Methods

Setting

Seventeen health care centers from seven regions across Spain were selected by the Spanish Back Pain Research Network to be invited to participate in this study, based on their involvement in previous research on LBP. Thirteen belonged to the Spanish National Health Service (SNHS) and four to not-for-profit foundations working for SNHS.

All the centers invited accepted. They included seven primary care centers and 10 specialty centers in rheumatology

EVIDENCE & METHODS

Context

Given the difficulties inherent in both the performance of RCTs and in the application of their results, the use of registries is gaining interest as a parallel and potentially powerful research tool.

Contribution

Using registry data on more than 4,000 Spanish patients, the authors found that several factors, including not having undergone previous surgery and shorter pain duration, were associated with a higher probability of improvement at 3 months relative to the baseline clinic visit.

Implication

While the observed associations are interesting, the real value of this paper lies in its relatively clear discussion of the strengths and limitations of the use of registry data in spinal research. It is important for readers to remember that registry data, with rare exceptions, are Level IV evidence.

—The Editors

(two), rehabilitation (two), neuroreflexotherapy (NRT) (three), and orthopedic surgery (three).

Subjects

Inclusion criteria were the following: seeking care in a participating center for LBP with or without leg pain, not caused by direct trauma or systemic diseases; not complying with criteria for immediate referral to surgery; and being able to read Spanish.

Pain not caused by systemic diseases was defined as pain in patients who had not been diagnosed with cancer, fibromyalgia, or inflammatory diseases (eg, rheumatoid arthritis or Bechterew's disease) and who did not show signs, which may suggest fibromyalgia (defined as diffuse pain with unexplained fatigue or sleep disturbances) or "red flags" for potential underlying systemic diseases, unless the appropriate diagnostic tests had ruled out this possibility. "Red flags" were defined as oncologic disease during the previous 5 years; constitutional symptoms such as unexplained weight loss, fever, chills; history of intravenous drug use; or immunocompromised host [7–10].

"Criteria for immediate referral to surgery" were defined as signs suggesting cauda equina syndrome (relevant or progressive paresia, loss of sphincter control, or saddle anesthesia), potential nerve root compression by disc herniation qualifying for surgery (disabling sciatic pain lasting 6 weeks or more, caused by nerve root compression from a disc protrusion or hernia, confirmed by magnetic

resonance imaging [MRI]), or symptomatic spinal stenosis for more than 3 months (defined as claudication unrelated to peripheral vascular disease with evidence of stenosis on MRI or computed tomography [CT] scans) [10–13]. Patients who had undergone unsuccessful spine surgery ("failed back surgery") and those with "red flags" in which appropriate test procedures had ruled out systemic diseases were included in the study.

Exclusion criteria were the following: treated or untreated central nervous system impairment and refusal to sign the informed consent.

The observational design of this study excluded potential ethical concerns regarding the sample size being too large. To analyze the influence of up to 40 variables on the clinical prognosis, sample size had to be large enough to ensure that at least 400 recruited patients would not improve [14]. Previous studies in SNHS have shown that the proportion of patients who show a clinically relevant improvement at 3 months for LBP, leg pain, and disability is approximately 80% to 85% and that losses to follow-up at 3 months are in the 5% to 10% range [15,16]. Hence, sample size for this study was established at 4,400.

Procedure

The study protocol was approved by the ethical committees of the participating entities.

Patients were recruited consecutively at the participating centers. All patients seeking care for LBP from physicians participating in the study were screened for inclusion and exclusion criteria. All patients complying with inclusion criteria were invited to participate. Physicians explained to eligible patients the importance of answering fully and accurately a series of questionnaires assessing their clinical status and complying with the follow-up visit for an assessment of their evolution. Patients were asked for an informed consent, which allowed the use of demographic and clinical data deriving from their care for the purpose of this study. All patients who signed it were included in the study. Following standard practice within SNHS, neither patients nor recruiting physicians received any compensation for their participation in this study.

At the design phase, the follow-up assessment was scheduled at 3 months for the following reasons: this timeframe was considered to be appropriate to assess the outcome of an episode of LBP; it ensures that all patients who would remain symptomatic at follow-up would be chronic [17]; for those patients who were already chronic on recruitment, the prognosis has been shown to be determined by changes in pain and disability during that period [18]; previous studies within SNHS have shown that losses to follow-up are minimal up to 3 months [15,19–21], start to rise at 6 months [22–24], and are substantial thereafter [16,25].

At both assessments, patients were asked to complete all the self-administered questionnaires in private, with no influence from health care personnel or other third parties. Once

completed, the questionnaires were collected and introduced into a database by auxiliary personnel independent from the treating physician. Participating in this study did not imply any changes to patients' clinical management, and physicians were instructed to manage their patients as usual. To make clinical decisions, the treating physicians had access to that information but were not able to alter the data.

Variables

Variables and measuring instruments selected for the current registry were based on those included in the only registry for LBP patients, which previously existed in SNHS. It had been originally developed for post-marketing surveillance of NRT and had shown to be reliable for gathering clinically meaningful data in routine practice [19,26–28].

At the first assessment, patients were asked to complete questionnaires on sex, age (date of birth), duration of the current pain episode (date of pain appearance), and employment status (classified as working, on sick leave, receiving disability compensation, student, housewife, unemployed, retired, or other, and grouping these categories at the analysis phase into “working;” “receiving financial assistance for LBP;” on sick leave or disabled for that reason; or “passive;” any other status). From September 8, 2009, the variable “academic level” (no studies, primary school, secondary school, university) was added to the questionnaire.

Severity of pain (both LBP and pain down to the leg [LP]) and degree of LBP-related disability are considered two of the main outcome measures for LBP patients [29]. At each assessment, patients' questionnaires included two separate 10-cm visual analog scales (VASs) for LBP and LP (for which 0=no pain and 10=worst possible pain) [30] and a validated Spanish version of the Roland-Morris questionnaire (RMQ) to measure LBP-related disability (for which 0=no disability and 24=worst possible disability) [20].

Recruiting physicians introduced the following information into the database: previous lumbar surgery (yes/no), whether the current episode was because of failed back surgery (yes/no), diagnostic procedures prescribed for the current episode (X-rays; CT scan; MRI; electromyography [EMG]; other, blood analyses, scintigraphy, and so forth), patients' radiological findings on imaging procedures performed for the current or previous episodes, as reported by radiologists (disc degeneration, facet joint degeneration, scoliosis, difference in leg length, spondylolysis, spondylolisthesis, spinal stenosis, annular tear, disc protrusion, disc herniation, lumbarization of S1, sacralization of L5, other radiological findings, no findings), and treatments that the patient underwent throughout the study (drugs: analgesics, nonsteroidal anti-inflammatory drugs, steroids, muscle relaxants, opioids; or other drugs: physiotherapy and rehabilitation, which were collapsed into a single category at the analysis phase and were defined as any form of exercise, heat,

cold, electrotherapy, or hands-on techniques, such as massage or mobilization; NRT intervention defined as the implantation of surgical material in specific areas of the skin, for up to 90 days [28]; surgery; other non-pharmacological treatments such as injections and spinal cord stimulation).

Psychological factors have shown to influence the clinical course and prognosis of LBP [18,31–36]. However, previous prospective studies have failed to identify the psychological variables, which have a relevant influence on symptoms or prognosis of LBP specifically in the Spanish cultural setting [22–24,37,38]. Therefore, no psychological variables could be included in the registry.

Analysis

The analysis was performed by a team of biostatisticians who were independent and had no contact with the clinicians involved in this study.

At the design phase of this study, it was hypothesized that subsequent LBP episodes in the same patient might not be independent events. Therefore, it was decided that only data from the first episode would be analyzed for each patient.

Absolute and relative frequencies were calculated for categorical variables. Values for continuous variables were described through their mean and standard deviation or median and percentiles 25 and 75, depending on whether their distribution was normal.

Improvements in pain and disability were defined as any reduction in the VAS or RMQ scores being higher than the minimal clinically important change (MCIC). Previous studies have established MCIC for pain and disability at 30% of their baseline score, with a minimum value of 1.5 for VAS and 2.5 for RMQ [15,39]. Roland-Morris questionnaire cannot be scored with decimals, so, in this study, improvements in pain and disability were defined as “clinically relevant” if ≥ 1.5 VAS points or ≥ 3 RMQ points, respectively. These definitions made it impossible, for patients with a baseline score below the corresponding cutoff point for a given variable, to show a clinically relevant improvement for that variable. For instance, it was impossible for a patient without leg pain at baseline to experience improvement in leg pain, and including this patient's data into the model exploring factors associated with a clinically relevant improvement in leg pain could skew results. In patients with baseline scores below MCIC, lack of improvement could only be identified in the event of worsening (ie, an increase in baseline score at discharge). Therefore, among patients with a baseline score below MCIC for a given variable, only those who had worsened at discharge were included in the analysis on factors associated with the prognosis of that variable.

Three separate multivariate predictive logistic regression models were developed to quantify the likelihood for a given patient to experience a clinically relevant improvement in LBP, LP, and disability. Improvement in LBP, leg pain, or disability were the dependent variables, and the

maximal models included sex, age (in years), baseline severity of LBP (VAS points), baseline severity of LP (VAS points), LBP-related disability at baseline (RMQ points), duration of the current episode (classified as acute, subacute, or chronic, with a cutoff limit for subacute and chronic at 14 and 90 days, respectively) [17,21], employment status (“passive,” which was the reference category; “working;” or “receiving financial assistance for LBP”), recruitment setting (primary care or specialized practice), history of lumbar surgery, diagnosis of failed back surgery syndrome, diagnostic tests undertaken at any moment during the study period (X-rays, CT scan, MRI, EMG, other), findings in imaging procedures undertaken during the study period or previous episodes (disc degeneration, facet joint degeneration, scoliosis, difference in leg length, spondylolysis, spondylolisthesis, spinal stenosis, annular tears, disc protrusion, disc herniation, lumbarization of S1, sacralization of L5, other findings, no findings), and treatments used during the study period (drugs such as analgesics, nonsteroidal anti-inflammatory drugs, steroids, muscle relaxants, opioids, other; physiotherapy/rehabilitation; NRT intervention; surgery; other). A nonautomatic backward strategy was followed, using the value $p > .05$ to eliminate variables from the model. Nomograms were drawn to illustrate the results of these models [14].

To assess the accuracy of the final model, both their discrimination and calibration were evaluated. Discrimination was assessed with the area under the receiver operating characteristic curve (AUC), whereas the Hosmer-Lemeshow test was used to assess calibration [14]. Bootstrapping was used to validate the selection of variables included in the model: 1,000 bootstrap samples were drawn [40]. Sample size was based on the number of observations without missing values [40,41]. The SWBOOT command available in Stata was used to validate the choice of variables [41]. The variables selected were displayed for each sample drawn, and the total number of times each variable was selected was counted [40].

To assess relevance of missing data, an imputation analysis was performed. Imputation of missing values was carried out by means of multiple imputation using chained equations [42]. A total of five imputed data sets were created. Selection of variables was performed separately in each data set. A variable was included in the combined model if it was selected as a predictor in at least two of the five imputed data sets [42,43]. Rubin rules were used to average the regression coefficients and standard errors of the predictors in the final model [44].

The SPSS v17 (SPSS Inc, Chicago, IL, USA), Stata v11.0 (StataCorp, College Station, TX, USA), and R v2.9 were used for statistical analysis.

Results

Data from 4,477 patients with a total of 5,506 episodes of LBP were gathered. Only data from the first pain episode

Table 1

Characteristics of patients who were recruited, diagnostic tests, and treatments used during the study

Variables	n=4,477
Sex (male)*	1,601 (35.8)
Age (y) [†]	53.8 (15.1)
Employment status*	
Passive	1,730 (38.6)
Receiving financial assistance for LBP	407 (9.1)
Working	2,340 (52.3)
Duration of the pain episode (d) [‡]	180 (90; 365)
Duration of the pain episode (d) categorized*	
Acute (<14 d)	78 (1.7)
Subacute (14–90 d)	1,495 (33.4)
Chronic (>90 d)	2,756 (61.6)
Missing	148 (3.3)
Baseline severity of LBP (VAS) [‡]	7.0 (5.0; 8.0)
Baseline severity of LP (VAS) (n=3,349) [‡]	7.0 (5.0; 8.0)
Baseline disability (RMQ) [†]	12.6 (5.8)
Previous lumbar surgery* (yes)	406 (9.1)
Failed back syndrome* (yes)	55 (1.2)
Diagnostic procedures during the episode*	
X-ray	1,213 (27.1)
MRI	1,385 (30.9)
CT scan	190 (4.2)
EMG	136 (3.0)
Other	44 (1.0)
Imaging findings*	
Disc degeneration	1,695 (37.9)
Facet joint degeneration	430 (9.6)
Scoliosis	204 (4.6)
Difference in leg length	21 (0.5)
Spondylolysis	65 (1.5)
Spondylolisthesis	155 (3.5)
Spinal stenosis	253 (5.7)
Disc protrusion	123 (2.7)
Disc herniation (extrusion)	1,059 (23.7)
Sacralization of L5	14 (0.3)
Lumbarization of S1	19 (0.4)
Other findings	17 (3.5)
No findings	150 (3.4)
Treatments	
Drugs*	2,749 (61.4)
Analgesics	2,622 (58.6)
NSAIDs	466 (10.4)
Steroids	935 (20.9)
Muscle relaxants	78 (1.7)
Opioids	3 (0.1)
Other	438 (9.8)
Non-pharmacological treatments*	
Physical therapy/rehabilitation	678 (15.1)
Neuroreflexotherapy	4,242 (94.8)
Surgery	50 (1.1)
Other	383 (8.6)

LBP, severity of low back pain; VAS, visual analog scale (range from better to worse; 0–10); LP, severity of pain referred down to the leg (in the 3,268 patients who had it); RMQ, Roland-Morris Questionnaire (range from better to worse; 0–24); MRI, magnetic resonance imaging; CT, computed tomography; EMG, electromyography; NSAIDs, nonsteroidal anti-inflammatory drugs.

* Frequency (%).

[†] Mean (standard deviation).

[‡] Median (P25; P75).

for each of the 4,477 patients were analyzed. Among the 4,477 patients included, 216 (4.8%) showed missing data for one or more variables: duration of pain (148 patients), baseline LBP severity (52 patients), or degree of LBP, LP, or disability at discharge (31 patients). However, there were no data on academic level for the 3,371 patients recruited before September 2009. Therefore, this variable was excluded from the analysis.

Among the 4,477 patients who were included in the analysis, most were women (64.2%) and the median duration of pain was 180 days. At baseline, the median value for pain severity was 7.0 VAS points (for both LBP and LP), and the mean degree of disability was 12.6 RMQ points (Table 1). The second assessment, which was planned at Day 90 after the first one, generally took place later (median, 102 days; interquartile range, 84.133). Only 30 patients (0.07%) were lost to follow-up, and their baseline characteristics were similar to those patients who attended it (data not shown).

Results from analyses including all the cases without missing values (“complete cases analysis”) were very similar to those from the imputation analysis (data not shown). Therefore, according to parsimony principle, only results from complete case analysis are shown [14].

Patients excluded from the analyses because their baseline scores were below the cutoff points, which allow considering potential improvement as “clinically relevant,” were 41 for LBP, 203 for LP, and 105 for disability.

Among the patients who were included in the corresponding models, 3,463 (78.6%) showed a clinically relevant improvement in LBP at discharge and the remaining 943 did not, whereas 3,102 patients (73.5%) showed a clinically relevant improvement in disability and 1,118 did not (Table 2). Among the 3,268 patients who reported pain referred LP when entering the study, 2,526 (75.2%) showed a clinically relevant improvement at discharge, and 833 did not (Table 2). Conversely, LP appeared during the study period in 91 patients who had not reported it at baseline. Among these patients, LP severity was mild; median (interquartile range) VAS score was 2.0 (1.0; 4.0).

The multivariate logistic regression model showed that the factors predicting a clinically relevant improvement in LBP at 3 months were the following (Table 3): clinical features at baseline (pain being acute, as opposed to subacute or chronic; higher severity of LBP; and lower severity of LP and disability); treatments received (not having undergone lumbar surgery and having been treated with NRT). Among the variables, which did not remain in the model, having been prescribed an MRI was the one with the highest frequency in bootstrapping validation (597) (Table 4). The assessment of the calibration of the model showed that the frequencies of expected and observed improvements were similar ($p=.202$). Discrimination of the model was poor, with an AUC of 0.648.

Factors predicting a clinically relevant improvement in LP at discharge were the following: clinical features at baseline (higher severity of LP and lower severity of LBP

and disability); not having been prescribed an EMG; and treatments received (not having undergone lumbar surgery and having been treated with NRT) (Table 4). Among the variables, which did not remain in the model, sex was the one with the highest frequency in bootstrapping validation (589) (Table 4). The assessment of the calibration of the model showed that the frequencies of expected and observed improvements were similar ($p=.156$). Discrimination of the model was poor, with an AUC of 0.655.

Factors predicting a clinically relevant improvement in disability at discharge were the following: clinical features at baseline (pain being acute, as opposed to subacute or chronic; higher degree of disability; and lower severity of LBP and LP); imaging findings (not showing disc degeneration); and treatments received (not having undergone lumbar surgery, not having received opioids, and having been treated with NRT and muscle relaxants) (Table 5). Among the variables, which did not remain in the model, spondylolisthesis was the one with the highest frequency in bootstrapping validation (432) (Table 5). The assessment of the calibration of the model showed that the frequencies of expected and observed improvements were similar ($p=.181$). Discrimination of the model was poor, with an AUC of 0.640.

Figs. 1–3 show the nomograms corresponding to the models on the improvement of LBP, LP, and disability, respectively.

Discussion

These results show that it is feasible to implement a registry for LBP patients treated in routine clinical practice. The registry included all patients seeking care for LBP who were treated following standard practice in routine clinical practice in the participating centers, gathered clinically relevant data through previously validated methods [20,29,30], and had less than 5% missing data and 1% losses to follow-up. Data from this registry made it possible to develop predictive models and nomograms to show the likelihood of improvement, factoring in patients' characteristics and clinical decisions (Tables 3–5; Figs. 1–3). The registry appears to not include all the information needed to build a valid prediction tool because models show a poor discrimination. However, their calibration is reasonable; AUC values are only slightly below 0.70, and results from the models were not affected by the recruitment setting, which suggests that the nomograms may be applicable in both primary and specialized care. Providing patients with data deriving from these nomograms may empower them for informed shared decision making.

A registry analysis is an observational study, and “association” does not necessarily imply “causality.” Therefore, results showing the association between a given variable and a better or worse prognosis should be interpreted cautiously, taking clinical plausibility into account. For

Table 2
Characteristics of patients in whom LBP, LP, and disability improved and did not improve

Variables	LBP			Leg pain			Disability	
	Showed a clinically relevant improvement (n=3,463)	Did not show any clinically relevant improvement (n=943)		Showed a clinically relevant improvement (n=2,526)	Did not show any clinically relevant improvement (n=833)		Showed a clinically relevant improvement (n=3,102)	Did not show any clinically relevant improvement (n=1,118)
Sex (male)*	1,211 (35.0)	357 (37.9)		866 (34.3)	278 (33.4)		1,110 (35.8)	399 (35.7)
Age (y) [†]	54.0 (15.1)	53.3 (15.1)		54.8 (14.7)	54.3 (14.5)		53.7 (15.1)	54.4 (14.9)
Employment status*								
Passive	1,350 (39.0)	353 (37.4)		1,044 (41.3)	322 (38.7)		1,187 (38.3)	444 (39.7)
Receiving financial assistance for LBP	301 (8.7)	103 (10.9)		218 (8.6)	95 (11.4)		303 (9.8)	97 (8.7)
Working	1,812 (52.3)	487 (51.6)		1,264 (50.0)	416 (49.9)		1,612 (51.8)	577 (51.6)
Duration of the pain episode (d) [‡]	180 (90; 365)	365 (90; 730)		180 (90; 450)	300 (90; 730)		180 (90; 365)	300 (90; 730)
Duration of the pain episode (d), categorized as*								
Acute (<14 d)	51 (1.5)	17 (1.8)		26 (1.1)	10 (1.2)		47 (1.6)	20 (1.9)
Subacute (14–90 d)	1,202 (36.1)	262 (28.2)		840 (34.4)	229 (28.1)		1,064 (35.5)	316 (29.2)
Chronic (>90 d)	2,077 (62.4)	649 (69.9)		1,574 (64.5)	576 (70.7)		1,885 (62.9)	745 (68.9)
Baseline LBP severity (VAS) [‡]	7.0 (6.0; 9.0)	7.0 (5.0; 8.0)		7.5 (6.0; 9.0)	7.0 (6.0; 8.0)		7.0 (5.5; 8.0)	7.0 (6.0; 8.1)
Baseline LP severity (VAS) [‡]	6.0 (1.0; 8.0)	5.0 (3.0; 8.0)		7.0 (5.0; 9.0)	6.0 (4.0; 8.0)		7.0 (5.0; 8.0)	7.0 (5.0; 8.0)
Baseline disability (RMQ) [‡]	12.6 (5.8)	13.2 (5.8)		13.1 (5.7)	14.0 (5.6)		13.6 (5.2)	12.3 (5.4)
Previous lumbar surgery*	290 (8.4)	113 (12.0)		237 (9.4)	116 (13.9)		262 (8.4)	132 (11.8)
Failed back surgery syndrome*	34 (1.0)	21 (2.2)		26 (1.0)	19 (2.3)		33 (1.1)	21 (1.9)
Diagnostic procedures undergone during the episode*								
X-ray	879 (25.4)	300 (31.8)		623 (24.7)	225 (27.0)		810 (26.1)	329 (29.4)
MRiv	1,031 (29.8)	324 (34.4)		781 (30.9)	308 (37.0)		971 (31.3)	358 (32.0)
CT scan	131 (3.8)	54 (5.7)		111 (4.4)	40 (4.8)		128 (4.1)	55 (4.9)
EMG	77 (2.2)	52 (5.5)		59 (2.3)	51 (6.1)		75 (2.4)	55 (4.9)
Other	35 (1.0)	7 (0.7)		30 (1.2)	4 (0.5)		30 (1.0)	13 (1.2)
Imaging findings*								
Disc degeneration	1,309 (37.8)	365 (38.7)		962 (38.1)	347 (41.7)		1,182 (38.1)	461 (41.2)
Facet joint degeneration	341 (9.8)	80 (8.5)		241 (9.5)	78 (9.4)		298 (9.6)	121 (10.8)
Scoliosis	162 (4.7)	38 (4.0)		109 (4.3)	34 (4.1)		131 (4.2)	63 (5.6)
Difference in leg length	14 (0.4)	4 (0.4)		11 (0.4)	3 (0.4)		13 (0.4)	4 (0.4)
Spondylolysis	49 (1.4)	14 (1.5)		32 (1.3)	11 (1.3)		39 (1.3)	23 (2.1)
Spondylolisthesis	119 (3.4)	33 (3.5)		83 (3.3)	36 (4.3)		103 (3.3)	49 (4.4)
Spinal stenosis	191 (5.5)	59 (6.3)		164 (6.5)	65 (7.8)		183 (5.9)	63 (5.6)
Disc protrusion	84 (2.4)	33 (3.5)		63 (2.5)	26 (3.1)		85 (2.7)	32 (2.9)
Disc herniation	798 (23.0)	243 (25.8)		629 (24.9)	233 (28.0)		766 (24.7)	271 (24.2)
Lumbarization of S1	11 (0.3)	6 (0.6)		8 (0.3)	3 (0.4)		10 (0.3)	8 (0.7)
Sacralization of L5	11 (0.3)	2 (0.2)		7 (0.3)	4 (0.5)		7 (0.2)	5 (0.4)
Other findings	132 (3.5)	40 (4.2)		83 (3.3)	41 (4.9)		120 (4.1)	48 (4.3)
No findings	95 (2.7)	45 (4.8)		50 (2.0)	19 (2.3)		83 (2.7)	42 (3.8)

Treatments		LBP, severity of low back pain; VAS, visual analog scale (range from better to worse; 0–10); LP, severity of pain referred down to the leg (in the 3,268 patients who had it); RMQ, Roland-Morris Questionnaire (range from better to worse; 0–24); MRI, magnetic resonance imaging; CT, computed tomography; NSAIDs, nonsteroidal anti-inflammatory drugs.					
Drugs*		* Frequency (%).					
		† Mean (standard deviation).					
		‡ Median (P25; P75).					
Analgesics	2,120 (61.2)	593 (62.9)	1,561 (61.8)	515 (61.8)	1,925 (62.1)	708 (63.3)	
NSAIDs	2,010 (58.0)	562 (59.6)	1,466 (58.0)	492 (59.1)	1,845 (59.5)	660 (59.0)	
Steroids	330 (9.5)	126 (13.4)	270 (10.7)	113 (13.6)	313 (10.1)	136 (12.2)	
Muscle relaxants	704 (20.3)	210 (22.3)	535 (21.2)	188 (22.6)	671 (21.6)	228 (20.4)	
Opioids	54 (1.6)	22 (2.3)	38 (1.5)	22 (2.6)	51 (1.6)	27 (2.4)	
Other	328 (9.5)	106 (11.2)	238 (9.4)	100 (12.0)	315 (10.2)	116 (10.4)	
Non-pharmacological treatments*							
Physiotherapy/rehabilitation	472 (13.6)	183 (19.4)	359 (14.2)	142 (17.1)	435 (14.0)	200 (17.9)	
Neuroreflexotherapy	3,386 (97.8)	834 (88.5)	2,486 (98.4)	765 (91.8)	3,038 (98.0)	1,003 (89.7)	
Surgery	36 (1.0)	12 (1.3)	31 (1.2)	16 (1.9)	35 (1.1)	15 (1.3)	
Others	296 (8.5)	83 (8.8)	200 (7.9)	88 (10.6)	271 (8.7)	106 (9.4)	

* Frequency (%).

† Mean (standard deviation).

‡ Median (P25; P75).

instance, it would be inappropriate to assume that variations in the odds of improvement associated with the use of any given form of treatment reflect “effects” from that treatment, either specific (ie, deriving from its mechanism of action) or unspecific (eg, deriving from the placebo effect or patients’ beliefs and expectations). Such variations could also reflect other factors, such as regression to the mean, natural history, or the association of the use of that form of treatment with any other variable that might influence patients’ or clinicians’ attitude toward that treatment. This should be taken into account especially when interpreting associations with a small predictive value, which are not supported by RCTs. Bearing this in mind, the association between the use of opioids and a poorer prognosis for disability might be interpreted as suggesting that these drugs are associated with a detrimental effect on this variable or the other way around; physicians may be more prone to prescribing opioids to patients in whom they perceive that disability is less likely to improve or is not the major clinical problem. It also might be interpreted that clinicians are more likely to prescribe muscle relaxants—instead of other drugs—to patients who they feel are more likely to do well (Table 5). The same caution should be applied to the interpretation of the prognostic value of undergoing an EMG; it is likely that clinicians are more prone to prescribing such a test to patients whose leg pain does not improve after initial treatment or to those in whom they suspect serious nerve compression; this makes more clinical sense than assuming that undergoing this test has a detrimental effect on the prognosis of leg pain (Table 4). Similarly, physicians’ interpretation of the relevance of images of disc degeneration in a given patient may be influenced by their perception of the severity of the patient’s disability (Table 5).

Nonetheless, registries can be useful to assess the prognostic value of procedures, which have previously shown to be effective through appropriate randomized clinical trials, in routine practice. In fact, the association between each treatment and the probability of improvement in pain and disability is generally consistent with the results from RCTs and systematic reviews on medication, NRT, and physiotherapy/rehabilitation (Tables 3–5) [19,24–28,45–50]. In general, the prognostic value of baseline clinical characteristics are also in line with what could be expected; a higher baseline value for a given variable (LBP, leg pain, or disability) leaves more room for its improvement, whereas the prognosis is worse for patients who are more severely affected (in terms of the other variables) and for those who are chronic, especially if they have already undergone lumbar surgery [10].

In this study, having undergone previous lumbar surgery was associated with a slightly worse prognosis for LBP, LP, and disability, whereas undergoing lumbar surgery during the follow-up period was not associated with any changes in the likelihood of improvement (Tables 3–5). However, this study is not appropriate for assessing the latter; on

Table 3

Predictive model for the improvement of LBP (n=4,239)*

Variables	Coefficients	p	Frequency in bootstrapping validation
Having been treated with neuroreflexotherapy	2.06	.000	1,000
Previous surgery	−0.38	.002	891
Baseline degree of disability (RMQ)	−0.04	.000	1,000
Baseline severity of LBP (VAS)	0.22	.000	1,000
Baseline severity of referred pain (VAS)	−0.04	.001	995
Chronicity			
Acute (reference category)			
Subacute	−0.88	.000	858
Chronic	−1.29	.009	1,000
Constant	−0.21	.511	

LBP, low back pain; RMQ, Roland-Morris Questionnaire (range from better to worse; 0–24); VAS, visual analog scale (range from better to worse; 0–10).

* The maximal model included sex, age, baseline severity of LBP, baseline severity of leg pain, baseline score for disability, chronicity, employment status, recruitment setting, history of lumbar surgery, diagnosis of failed back surgery syndrome, diagnostic tests undertaken at any moment during the study period, findings in imaging procedures undertaken during the study period or previous episodes, and treatments used during the study period.

the one hand, the potential prognostic value of surgery may appear beyond 3 months. On the other hand, this study excluded patients who at baseline showed well-established surgical criteria (ie, cauda equina syndrome or symptomatic disc herniation or spinal stenosis despite conservative treatment) [11–13]. As a result, only 50 patients (1.1%) underwent surgery during the study period (Table 1), which made it difficult for any potential prognostic value of

Table 4

Predictive model for improvement of LP* (n=3,359)

Variables	Coefficients	p	Frequency in bootstrapping validation
Having been treated with neuroreflexotherapy	1.47	.000	1,000
Previous surgery	−0.49	.000	940
Baseline degree of disability (RMQ)	−0.05	.000	1,000
EMG	−0.52	.021	569
Baseline severity of low back pain (VAS)	−0.07	.013	890
Baseline severity of referred pain (VAS)	0.21	.000	1,000
Constant	−0.48	.063	

LP, pain down to the leg; RMQ, Roland-Morris Questionnaire (range from better to worse; 0–24); EMG, electromyography; VAS, visual analog scale (range from better to worse; 0–10).

* The maximal model included sex, age, baseline severity of low back pain, baseline severity of leg pain, baseline score for disability, chronicity, employment status, recruitment setting, history of lumbar surgery, diagnosis of failed back surgery syndrome, diagnostic tests undertaken at any moment during the study period, findings in imaging procedures undertaken during the study period or previous episodes, and treatments used during the study period.

Table 5

Predictive model for improvement of disability* (n=4,076)

Variables	Coefficients	p	Frequency in bootstrapping validation
Having been treated with neuroreflexotherapy	2.25	.000	1,000
Disc degeneration diagnosis	−0.19	.015	512
Previous surgery	−0.45	.000	964
Muscle relaxants	0.24	.012	720
Opioids	−0.60	.016	682
Baseline degree of disability (RMQ)	0.06	.000	1,000
Baseline severity of low back pain (VAS)	−0.05	.024	706
Baseline severity of referred pain (VAS)	−0.03	.022	704
Chronicity			
Acute (reference category)			
Subacute	−0.85	.012	837
Chronic	−1.34	.000	1,000
Constant	−0.14	.667	

RMQ, Roland-Morris Questionnaire (range from better to worse; 0–24); VAS, visual analog scale (range from better to worse; 0–10).

* The maximal model included sex, age, baseline severity of low back pain, baseline severity of leg pain, baseline score for disability, chronicity, employment status, recruitment setting, history of lumbar surgery, diagnosis of failed back surgery syndrome, diagnostic tests undertaken at any moment during the study period, findings in imaging procedures undertaken during the study period or previous episodes, and treatments used during the study period.

surgery to be statistically significant. Moreover, these 50 patients may have included those in whom surgical criteria appeared during follow-up and also those without such criteria (eg, only axial pain), for whom the evidence on the effectiveness of surgery is controversial [51].

This study recruited LBP patients who should be managed according to recommendations from evidence-based clinical guidelines for common LBP, that is, patients with neither surgical criteria nor “red flags” for systemic diseases [10] and those with “red flags” in whom appropriate tests had ruled out systemic diseases. According to the evidence available [52–58], this may account for the fact that, in general, performing diagnostic tests [52] or establishing specific diagnoses based on imaging (eg, scoliosis, disc protrusion or herniation, facet joint degeneration, spinal stenosis, and so on) [52–58] did not show a prognostic value for predicting the clinical evolution of pain and disability at 3 months (Tables 3–5).

In Spain, workers’ compensation benefits can reach up to 100% of the salary, and receiving financial benefits has shown to be associated with delayed recovery [59–63]. Contrarily to what was expected at the design phase of this study, results from the regression analyses showed that receiving economic compensation for LBP had no prognostic value for predicting the clinical evolution (Tables 3–5). It is impossible to rule out that the cultural setting where this study took place accounts for this finding because, to the best of the authors’ knowledge, no previous studies have analyzed whether this variable is predictive specifically

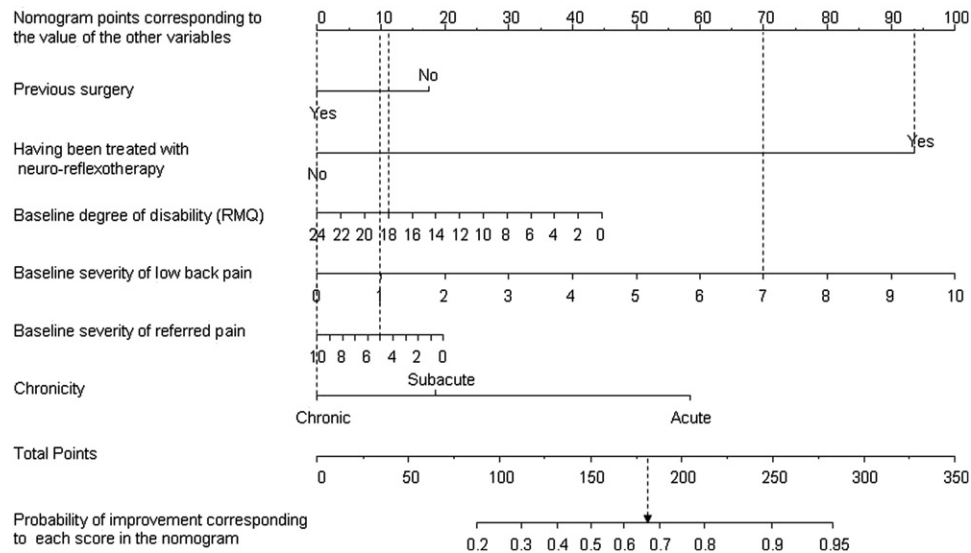


Fig. 1. Nomogram corresponding to the models on the improvement of low back pain. RMQ, Roland-Morris questionnaire.

among Southern European patients. However, this finding may also be explained by the fact that this study focused on the clinical evolution at 3 months, whereas workers' compensation benefits in Spain last for up to 18 months, which raises the possibility that the predictive value of this variable appears after 3 months. Moreover, most patients included in this study were chronic and only 9.1% were receiving such benefits (Table 2). Therefore, future studies with longer follow-up periods and including larger samples, a higher proportion of acute patients and more subjects receiving workers' compensation benefits, should assess the predictive value of this variable in this cultural context.

Previous studies and systematic reviews on prognosis factors for LBP vary substantially in design and setting,

and methodological shortcomings have led to uncertainty about the reliability of the prognostic factors, which have been suggested [58]. In fact, many original studies on prognostic factors are reanalyses of RCTs with relatively small samples, aiming at identifying subsets of patients who respond better to some form of treatment [18,64]. This makes it inappropriate to compare their results with those from this study, in which a large population recruited in routine clinical practice and receiving a broad array of treatments was analyzed.

This study and the registry from which it derives have a number of limitations. Losses to follow-up beyond 3 months are unknown and should be assessed in further studies. Longer follow-up periods may be more appropriate to

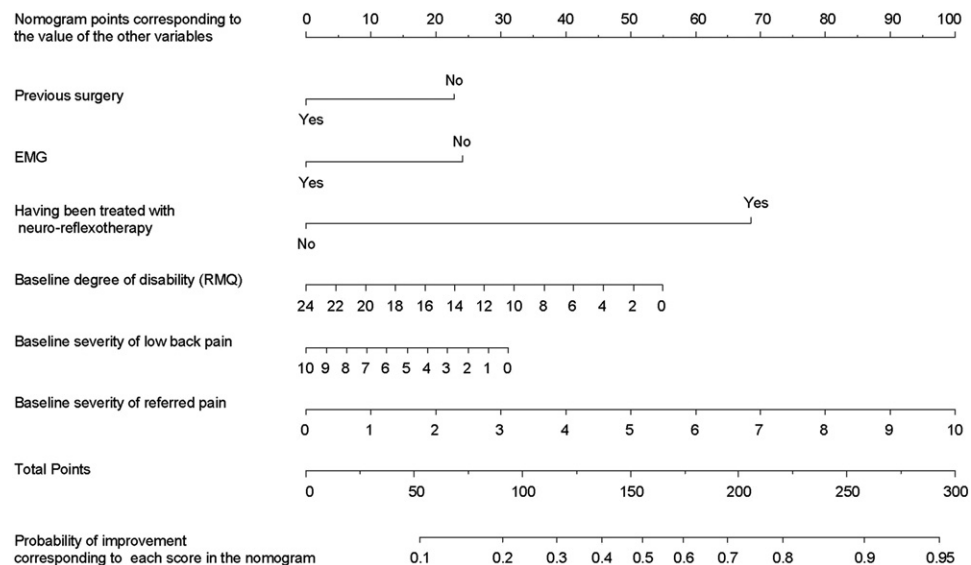


Fig. 2. Nomogram corresponding to the models on the improvement of pain down to the leg. EMG, electromyography; RMQ, Roland-Morris questionnaire.

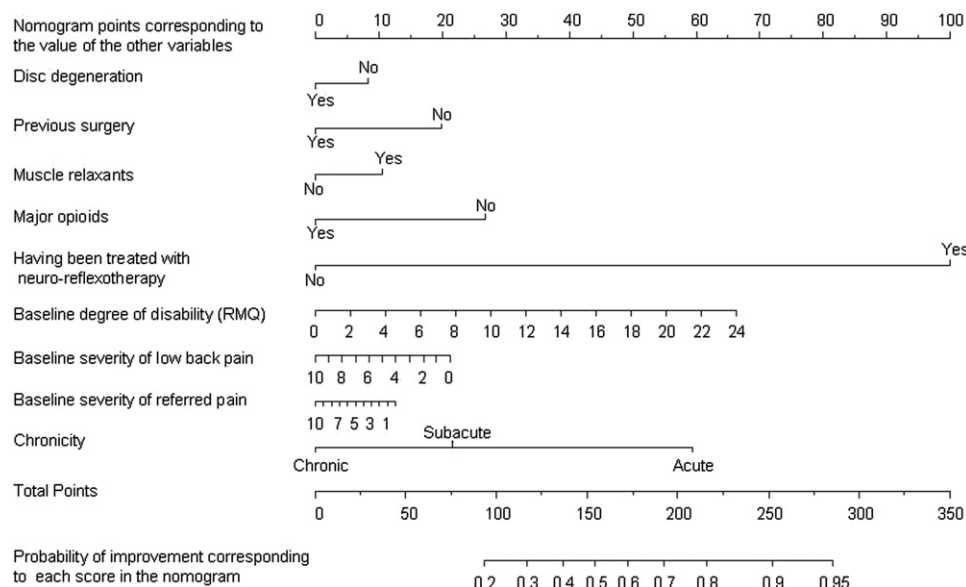


Fig. 3. Nomogram corresponding to the models on the improvement of disability. RMQ, Roland-Morris questionnaire.

assess the predictive values of some variables, such as receiving workers' compensation benefits or having undergone surgery for the current episode. The current analysis focused on the clinical prognosis of an LBP episode and did not analyze data on the factors associated with relapses or with adverse events attributed to the different forms of treatment, which also require longer follow-up periods and larger samples.

Spanish National Health Service is a tax-funded public organization in which all health care services are provided free to every resident in Spain, and digital clinical history is implemented within routine practice across primary and specialized care. Within SNHS, there are no incentives to encourage physicians to refer patients to specialists or to recommend specific diagnostic tests or treatments, and neither patients nor physicians receive any economic compensation for participating in studies. Despite this, this study had less than 5% missing data and losses to follow-up. This is a common feature to most studies conducted in this setting with less than 6 months' follow-up [15,19–21] and probably reflects social and cultural characteristics of Spanish patients. It may be more difficult to implement registries in settings with different cultural backgrounds, a fee-per-service health system, or where the clinical community is less familiar with e-health.

Health care centers that participated in this study were selected because of their previous involvement in LBP research and are not a random sample of health care centers within SNHS. As a result, participating clinicians may be more evidence based in their practice or more positively predisposed than average for implementing a registry on LBP. However, these centers represented a feasible and adequate platform for pilot testing this registry in routine practice; LBP-related clinical practice is roughly consistent

across SNHS [16], and it generally follows recommendations from evidence-based clinical guidelines [16].

The vast majority of patients included in this study had a pain episode that lasted over 14 days (Table 1), possibly because patients were recruited both in primary and specialized care, where subacute and chronic patients are more prevalent. Moreover, patients with pain for only a few days are less likely to feel motivated to commit to returning for follow-up 3 months later, which suggests that most patients who refused to participate in this study, or were lost to follow-up, may have been acute. As a result, predictive models developed in this study may be less valid at the very early stage of the pain episode, and this should be assessed in further studies. Nonetheless, the sample included in this study can be seen as fairly representative of subacute and chronic LBP patients seeking care in SNHS, who represent the most difficult clinical challenge and generate the largest proportion of social costs. Moreover, the number of acute patients included in this study was high enough for the multivariate predictive logistic regression models to quantify variations in the odds for improvements in pain and disability associated with the chronicity status (Tables 3–5; Figs. 1–3).

Procedures considered as “alternative medicine,” such as acupuncture or homeopathy, were not included in this registry. However, there is no evidence supporting the notion that undergoing these procedures has any predictive value on the evolution of LBP [48,49,63,65–69]. Moreover, because these procedures are not covered by SNHS, their use among Spanish patients is infrequent. Therefore, it is not likely that having excluded them from the registry is a major concern with regard to the validity of the results of this study within SNHS.

The psychological variables influencing the prognosis of LBP patients specifically in the Spanish cultural setting are

yet to be identified; to date, only the influence of fear avoidance beliefs and catastrophizing have been assessed and shown to be clinically irrelevant or null [22–24,37,38]. As a result, the registry currently does not include any psychological variables. However, pain is a sensory and emotional experience, which is influenced by psychological factors [18,31–36]. Therefore, once psychological variables with a prognostic value in this setting are identified, they should be added to this registry. Nonetheless, despite the absence of psychological variables, the predictive models make it possible to quantify the likelihood of clinical improvement for LBP patients treated in routine practice within SNHS (Tables 3–5; Figs. 1–3).

In conclusion, this study shows that it is possible to use data gathered through a registry in routine clinical practice. Future studies should assess the generalizability of the predictive models built using the data gathered by this registry.

References

- [1] Carragee EJ, Deyo RA, Kovacs FM, et al. Clinical research. Is the spine field a mine field? *Spine* 2009;34:423–30.
- [2] Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med* 2009;151:203–6.
- [3] Dreyer NA, Garner S. Registries for robust evidence. *JAMA* 2009;302:790–1.
- [4] Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA* 2008;299:656–64.
- [5] Institute of Medicine. Committee on comparative effectiveness research prioritization, initial national priorities for comparative effectiveness research. Washington, DC: National Academies Press, 2009. Available at: www.iom.edu/CMS/3809/63608/71025.aspx.
- [6] Schoene ML. Are patient registries ready for prime time? *The Back Letter* 2009;11:121–6.
- [7] Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992;268:760–5.
- [8] Slipman CW, Patel RK, Botwin K, et al. Epidemiology of spine tumors presenting to musculoskeletal physiatrists. *Arch Phys Med Rehabil* 2003;84:492–5.
- [9] van den Hoogen HMM, Koes BW, van Eijk JThM, Bouter LM. On the accuracy of history, physical examination and erythrocyte sedimentation rate in diagnosing low back pain in general practice. A criteria-based review of the literature. *Spine* 1995;20:318–27.
- [10] Waddell G. *The back pain revolution*. London: Churchill-Livingstone, 1998.
- [11] Peul WC, van de Hout W, Brand R, et al. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. *BMJ* 2008;336:1355–8. originally published online 23 May 2008; doi: 10.1136/bmj.a143.
- [12] Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation. Four-year results for the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008;33:2789–800.
- [13] Kovacs FM, Urrutia G, Alarcón JD. Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. *Spine* 2011;36:E1335–51.
- [14] Harrell FE. *Regression modeling strategies with applications to linear models, logistics regression, and survival analysis*. New-York: Springer, 2001.
- [15] Kovacs FM, Abraira V, Royuela A, et al. Spanish Back Pain Research Network. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007;32:2915–20.
- [16] Kovacs FM, Fernandez C, Cordero A, et al, and the Spanish Back Pain Research Network. Non-specific low back pain in primary care in the Spanish National Health Service: a prospective study on clinical outcomes and determinants of management. *BMC Health Serv Res* 2006;6:57.
- [17] Merskey H, Bogduk N. Description of chronic pain syndromes and definitions of pain terms. In: *Classification of chronic pain*. 2nd ed. Seattle: IASP Press, 1994.
- [18] Heymans MW, van Buuren S, Knol DL, et al. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. *Spine J* 2010;10:847–56.
- [19] Kovacs FM, Abraira V, Muriel A, et al. Prognostic factors for neuro-reflexotherapy in the treatment of subacute and chronic neck and back pain. A study of predictors of clinical outcome in routine practice of the Spanish National Health Service. *Spine* 2007;32:1621–8.
- [20] Kovacs FM, Llobera J, Gil del Real MT, et al. Validation of the Spanish version of the Roland Morris Questionnaire. *Spine* 2002;27:538–42.
- [21] Kovacs FM, Abraira V, Zamora J, Fernández C, and the Spanish Back Pain Research Network. The transition from acute to subacute and chronic low back pain. A study based on determinants of quality of life and prediction of chronic disability. *Spine* 2005;30:1786–92.
- [22] Kovacs FM, Muriel A, Abraira V, et al. The influence of fear avoidance beliefs on disability and quality of life is sparse in Spanish low back pain patients. *Spine* 2005;30:E676–82.
- [23] Kovacs F, Abraira V, Santos S, et al. A comparison of two short education programs for improving low back pain-related disability in the elderly. A cluster randomized controlled trial. *Spine* 2007;32:1053–9.
- [24] Albadalejo C, Kovacs FM, Royuela A, et al. The efficacy of a short education program and a short physiotherapy program for treating low back pain in primary care. A cluster randomized trial. *Spine* 2010;35:483–96.
- [25] Kovacs FM, Llobera J, Abraira V, et al, and the KAP Group. Effectiveness and cost effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice. A cluster randomized, controlled trial. *Spine* 2002;27:1149–59.
- [26] Grupo PINS (Pilotaje de la Intervención NRT en el Sistema Nacional de Salud). [Neuroreflexotherapy intervention for treating common neck and back pain within the Spanish National Health Service. Results from a pilot test]. *Gac Sanit* 2004;18:275–86.
- [27] Corcoll J, Orfila J, Tobajas P, Alegre L. Implementation of neuroreflexotherapy for subacute and chronic neck and back pain within the Spanish public health system: audit results after one year. *Health Policy* 2006;79:345–57.
- [28] Urrutia G, Burton A, Morral A, et al. Neuroreflexotherapy for non-specific low back pain. *Cochrane Database Syst Rev* 2004;(Issue 2): CD003009. <http://dx.doi.org/10.1002/14651858.CD003009.pub2>.
- [29] Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998;23:2003–13.
- [30] Huskinson EC. Measurement of pain. *Lancet* 1974;2:1127–31.
- [31] Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 2002;27:E109–20.
- [32] Smeets RJ, Vlaeyen J, Kester AD, Knottnerus JA. Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *J Pain* 2006;7:261–71.
- [33] Smeets RJ, Maher CG, Nicholas MK, et al. Do psychological characteristics predict response to exercise and advice for subacute low back pain? *Arthritis Rheum* 2009;61:1202–9.
- [34] Spinoven P, Ter Kuile M, Kole-Snijders AM, et al. Catastrophizing and internal pain control as mediators of outcome in the multidisciplinary treatment of chronic low back pain. *Eur J Pain* 2004;8:211–9.
- [35] Truchon M, Côté D. Predictive validity of the Chronic Pain Coping Inventory in subacute low back pain. *Pain* 2005;116:205–12.

- [36] Wand BM, McAuley JH, Marston L, De Souza LH. Predicting outcome in acute low back pain using different models of patient profiling. *Spine* 2009;34:1970–5.
- [37] Kovacs F, Abraira V, Cano A, et al. Spanish Back Pain Research Network. Fear avoidance beliefs do not influence disability and quality of life in Spanish elderly subjects with low back pain. *Spine* 2007;32:2133–8.
- [38] Kovacs FM, Noguera J, Abraira V, et al. The influence of psychological factors on low back pain-related disability in community-dwelling older persons. *Pain Med* 2008;9:871–80.
- [39] Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008;33:90–4.
- [40] Beyene J, Atenafu E, Hamid J, et al. Determining relative importance of variables in developing and validating predictive models. *BMC Med Res Methodol* 2009;9:64.
- [41] Garrett JM. SWBOOT: Stata module to bootstrap stepwise linear or logistic regression models. Available at: <http://ideas.repec.org/c/boc/bocode/s415601.html#refs>, last accessed February 23, 2012.
- [42] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377–99.
- [43] Schellingerhout JM, Heymans MW, Verhagen AP, et al. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine* 2010;35:E827–35.
- [44] Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.
- [45] Roelofs PD, Deyo RA, Koes BW, et al. Nonsteroidal anti-inflammatory drugs for low back pain: an updated Cochrane review. *Spine* 2008;33:1766–74.
- [46] Machado LA, Kamper SJ, Herbert RD, et al. Analgesic effects of treatments for non-specific low back pain: a meta-analysis of placebo-controlled randomized trials. *Rheumatology* 2009;48:520–7.
- [47] van Tulder MW, Touray T, Furlan AD, et al. Cochrane Back Review Group. Muscle relaxants for nonspecific low back pain: a systematic review within the framework of the Cochrane collaboration. *Spine* 2003;28:1978–92.
- [48] van Tulder MW, Koes B, Malmivaara A. Outcome of non-invasive treatment modalities on back pain: an evidence-based review. *Eur Spine J* 2006;15(Suppl 1):S64–81.
- [49] Chou R, Huffman LH. Medications for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:505–14.
- [50] Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492–504.
- [51] Chou R, Baisden J, Carragee EJ, et al. Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine* 2009;34:1094–109.
- [52] Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet* 2009;373:463–72.
- [53] Weinstein SL, Dolan LD, Spratt KF, et al. Health and function of patients with untreated idiopathic scoliosis. A 50-year natural history study. *JAMA* 2003;289:559–67.
- [54] Waris E, Eskelin J, Hermunen H, et al. Disc degeneration in low back pain. A 17-year follow-up study using magnetic resonance imaging. *Spine* 2007;32:681–4.
- [55] Carragee EJ, Alamin TF, Carragee JM. Low-pressure positive discography in subjects asymptomatic of significant low back pain illness. *Spine* 2006;31:505–9.
- [56] Modic MT, Obuchowski NA, Jeffrey S, et al. Acute low back pain and radiculopathy: MR imaging findings and their prognostic role and effect on outcome. *Radiology* 2005;237:597–604.
- [57] Kanayama M, Togawa D, Takahashi C, et al. Cross-sectional magnetic resonance imaging study of lumbar disc degeneration in 200 healthy individuals. *J Neurosurg Spine* 2009;11:501–7.
- [58] Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586–97.
- [59] Rainville J, Sobel JB, Hartigan C, Wright A. The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. *Spine (Phila Pa 1976)* 1997;22:2016–24.
- [60] Scuderi GJ, Sherman AL, Brusovanik GV, et al. Symptomatic cervical disc herniation following a motor vehicle collision: return to work comparative study of workers' compensation versus personal injury insurance status. *Spine J* 2005;5:639–44.
- [61] Rasmussen C, Leboeuf-Yde C, Hestbaek L, Manniche C. Poor outcome in patients with spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in the secondary care sector. *Scand J Rheumatol* 2008;37:462–8.
- [62] Atlas SJ, Chang Y, Kammann E, et al. Long-term disability and return to work among patients who have a herniated lumbar disc: the effect of disability compensation. *J Bone Joint Surg Am* 2000;82:4–15.
- [63] Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA* 2010;303:1295–302.
- [64] Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results—guidance for future prognosis reviews. *J Clin Epidemiol* 2009;62:781–96.
- [65] Cherkin DC, Sherman KJ, Ayins AL, et al. Simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* 2009;169:858–66.
- [66] Santaguida PL, Gross A, Busse J, et al. Complementary and alternative medicine in back pain utilization report. *Evid Rep Technol Assess (Full Rep)* 2009;177:1–221.
- [67] Furlan AD, van Tulder M, Cherkin D, et al. Acupuncture and dry-needling for low back pain: an updated systematic review within the framework of the Cochrane collaboration. *Spine* 2005;30:944–63.
- [68] Yuan J, Purepong N, Kerr DP, et al. Effectiveness of acupuncture for low back pain: a systematic review. *Spine* 2008;33:E887–900.
- [69] Weiner DK, Ernst E. Complementary and alternative approaches to the treatment of persistent musculoskeletal pain. *Clin J Pain* 2004;20:244–55.

ARTÍCULO 2:

**Predicting outcomes of neuroreflexotherapy in patients
with subacute or chronic neck or low back pain**

Justificación metodológica

En este segundo artículo, se desarrollan tres modelos predictivos para la mejoría en la intensidad del dolor lumbar o cervical, intensidad del dolor irradiado (a la pierna o ciática en caso del dolor lumbar, y al brazo en caso del dolor cervical) y discapacidad, definida como el grado de restricción de las actividades cotidianas que impone el dolor lumbar o cervical, en pacientes que han recibido una intervención neurorreflejojoterápica (NRT). El pronóstico se realiza para los tres meses siguientes a la intervención NRT.

Los datos de los participantes se recogieron gracias a un registro prospectivo que comenzó en el año 2002 y que contiene datos de pacientes que han sido intervenidos con la NRT. Para poder utilizar los datos de estos pacientes, previamente fueron informados y firmaron el correspondiente consentimiento informado.

La NRT es un tratamiento para las dolencias del cuello y la espalda. Consiste en implantar superficialmente en la piel (en los denominados “puntos gatillo” o *trigger points* de la espalda y en los puntos dolorosos referidos del pabellón auricular) un material quirúrgico, sobre fibras nerviosas que al ser estimuladas desencadenan efectos que contrarrestan los mecanismos que mantienen el dolor, la inflamación y la contractura muscular en los casos de dolor de cuello y espalda de carácter subagudo y crónico. La eficacia de este procedimiento se ha demostrado en diversos ensayos clínicos (52–54). Es además el único tratamiento para la lumbalgia que ha demostrado ser coste/efectivo en el Sistema Nacional de Salud español (55–57), y cuyas condiciones de aplicación están definidas y evaluadas (58,59).

Las grapas se implantan superficialmente en la piel, a una profundidad de menos de 2 mm. En la literatura, no se han informado hasta la fecha problemas en cuanto a dolor o cicatrización asociados con el procedimiento. La NRT se realiza sin anestesia y la implantación lleva aproximadamente 60 minutos. Las grapas permanecen implantadas hasta 90 días en la espalda y hasta 20 días en el pabellón auricular. En los casos en los que es necesario, la intervención puede repetirse.

La NRT puede confundirse con la acupuntura ya que ambas utilizan dispositivos de punción. Sin embargo, en ambas técnicas se estimulan diferentes zonas de la piel.

La NRT está indicada en pacientes en los que el dolor persiste tras 14 o más días de tratamiento farmacológico, con una intensidad de al menos 3 puntos en una escala analógica visual (cuyo intervalo oscila entre 0, ausencia de dolor, a 10, máximo dolor imaginable).

Siguiendo la línea argumental de esta tesis, en este segundo trabajo, se han incorporado nuevos aspectos derivados de los avances en el campo de los modelos pronósticos y su validación.

En la definición de las variables resultado u *outcomes*, se ha seguido la misma metodología que en el artículo anterior, siguiendo la línea de los cambios mínimos clínicamente relevantes y añadiendo los puntos de corte correspondientes a pacientes con cervicalgia en cuanto a intensidad de dolor y discapacidad (60).[‡]La inclusión de pacientes con una intensidad de dolor o discapacidad basales por debajo del punto de corte para el cambio mínimo clínicamente relevante (por tanto, sin posibilidad de mejoría) pero que hubieran empeorado a los tres meses, no parecía un criterio lo suficientemente sólido como para repetirse en este segundo estudio. En consecuencia, solamente se incluyeron aquellos pacientes con una intensidad de dolor (lumbar o cervical) y un nivel de discapacidad por encima del punto de corte establecido como clínicamente relevante para lumbalgia y cervicalgia, respectivamente.

Con relación a la selección de las variables para conseguir un modelo final o estrategia de modelización, no hay un consenso sobre el mejor método a desarrollar. La recomendación general es mantener en el modelo aquellos factores de riesgo que tienen credibilidad clínica y se encuentran bien establecidos en la literatura (20), así como evitar estrategias de selección de variables basadas en los valores-p (6,20,33). En este trabajo, no se ha llevado a cabo estrategia de modelización, puesto que el objetivo era obtener un pronóstico con la mayor exactitud posible, y para ello, se

[‡]En el Apéndice III se adjunta la correspondiente publicación.

incluyeron y mantuvieron en el modelo todas aquellas variables que cumplieran estos criterios. Además, dado el tamaño muestral del registro con el que se llevó a cabo el análisis, la posible falta de potencia y la amenaza de saturación del modelo, no fueron un problema a la hora de especificar el número de variables que entraron en él.

La importancia de considerar los datos faltantes en los análisis de resultados y la imputación múltiple, se ha incrementado en los últimos años (61). Aunque inicialmente la recomendación sobre el número de imputaciones oscilaba entre 3 y 5 (12,62), más tarde esa recomendación aumentó hasta 10 imputaciones (63). Por tanto, en este trabajo, se decidió llevar a cabo 10 imputaciones. Por otro lado, como no se estableció estrategia de modelización, la combinación de los resultados entre los 10 conjuntos de datos pudo llevarse a cabo de manera automática.

Otro de los puntos metodológicos que se presentan en este artículo, es la evaluación de la calibración incorporando técnicas de *bootstrapping* para construir un gráfico que tuviera en cuenta el sobre-ajuste del modelo e incorporara la información de las 10 imputaciones llevadas a cabo en la fase de análisis.

En la evaluación gráfica de la calibración se representan las probabilidades predichas por el modelo en el eje de abscisas y los resultados observados en el eje de ordenadas. La predicción perfecta se encuentra en la línea de los 45°, y la pendiente de la recta entre observaciones vs predicciones proporciona una medida de la calibración del modelo. En el caso de la regresión lineal, el gráfico de calibración, es simplemente un gráfico de dispersión. En el caso de resultados binarios, el gráfico sólo cuenta con los valores 0 y 1 en el eje Y. Las técnicas de suavizado o *smoothing* se pueden utilizar para estimar las probabilidades observadas del *outcome* en relación con las predichas.

La representación gráfica entre frecuencias observadas (O) y probabilidades predichas (P) se denomina “Curva aparente” (“*Apparent curve*”). Si O y P coinciden a lo largo de todo el rango de probabilidades, el gráfico mostrará una línea de 45° o “Curva perfecta” (“*Ideal curve*”). La ordenada en el origen de la curva, γ_0 , es igual a 0 y la pendiente, γ_1 , igual a 1. Valores en γ_0 distintos de 0 indicarían la medida en que las predicciones son sistemáticamente demasiado bajas o demasiado altas; y valores en γ_1 inferiores a 1 indicarían que el modelo está sobre-ajustado. La ordenada en el origen y

la pendiente de la curva de calibración se pueden estimar mediante una regresión logística con una sola variable independiente, el predictor lineal (6,23,64):

$$\log \text{odds}(\text{outcome}) = \gamma_0 + \gamma_1 * \text{predictor lineal}$$

Es posible extender el gráfico de calibración a un gráfico de validación, añadiendo una “Curva con el sesgo corregido” (“Bias-corrected curve”) obtenida mediante *bootstrapping*. La validación por *bootstrap*, fue descrita en detalle por Efron y Gong (65). Consiste en desarrollar el modelo con todos los sujetos, para después re-analizarlo en repetidas muestras aleatorias del conjunto de datos. Este método de validación interna evalúa la precisión con la que los modelos predecirán el *outcome* en una nueva muestra de pacientes similar a la que se desarrolló el modelo. De esta manera, se valora si el modelo se encuentra sesgado o no por sobre-ajuste. Aplicando la técnica de *bootstrapping* a la generación de una “Curva con el sesgo corregido”, el algoritmo estimará sucesivamente los valores de γ_0 y γ_1 en diferentes re-muestras *bootstrap* y dibujará la “Curva con el sesgo corregido” promediando los resultados (6).

Así, la “Curva perfecta” representa la calibración perfecta, la “Curva aparente” representa la calibración de los datos originales, y la “Curva con el sesgo corregido” representa la calibración del modelo corregida por *bootstrapping*.

De esta manera, la proximidad entre la “Curva con el sesgo corregido” a la “Curva perfecta”, demuestra que el modelo se encuentra bien calibrado. La proximidad entre la “Curva con el sesgo corregido” con la “Curva aparente”, señalará que el sobre-ajuste es mínimo. Aunque la curva de calibración del modelo original se acercaría a lo ideal, la “Curva con el sesgo corregido” es la representación más válida de la posible relación que podría observarse en otros conjuntos de datos similares, aunque se encuentre penalizada por el hecho de que los modelos se hayan desarrollado y validado sobre los mismos pacientes.

Hasta la fecha, en el conocimiento de la doctoranda, no se ha publicado nada acerca de este procedimiento en el caso de diversos conjuntos de datos creados a partir de la imputación múltiple. En esta Tesis, se desarrolla un gráfico de calibración a partir de varios conjuntos de datos consecuencia de una imputación múltiple.

De cada una de las imputaciones, se estimarán sus correspondientes γ_0 y γ_1 , obteniendo un valor de γ_0 y γ_1 en cada caso (tantos γ_0 y γ_1 como imputaciones se llevaron a cabo). Promediando los valores de γ_0 y γ_1 para cada modelo (Mejoría del dolor local, Mejoría del dolor irradiado y Mejoría de la discapacidad), se obtienen los valores globales para poder construir un gráfico de calibración global correspondiente a cada uno de ellos.

La evaluación de la discriminación se llevó a cabo mediante el *índice D_{xy} de Somer*. Un índice ampliamente utilizado para evaluar la discriminación es el *índice c* , o índice de concordancia (64). En el contexto de los modelos pronósticos con resultados binarios, se define como la proporción de parejas de pacientes u observaciones, uno que presente el evento de interés y otro que no, en las que el paciente con el evento tiene una mayor probabilidad predicha por el modelo. Esencialmente es lo mismo que el estadístico de Wilcoxon-Mann-Whitney para comparación de predicciones entre dos grupos, y es idéntico al área bajo la curva ROC. El índice *D_{xy} de Somer* es una medida de correlación de rangos entre las probabilidades predichas y las respuestas observadas. D_{xy} se relaciona con el *índice c* mediante la ecuación $D_{xy}=2(c-0,5)$ (6). Para cuantificar el optimismo del modelo ajustado, en esta tesis se ha estimado el *índice D_{xy}* mediante réplicas *bootstrap*.

Por tanto, los objetivos de este segundo artículo son: aplicar las técnicas de imputación múltiple de datos faltantes, desarrollar modelos válidos que predigan de un modo fiable la probabilidad de mejorar clínicamente la intensidad del dolor local lumbar y cervical, dolor irradiado a la pierna o al brazo y discapacidad asociada al dolor lumbar y cervical en pacientes a los que se les ha realizado una intervención NRT, y desarrollar un gráfico de calibración a partir de los múltiples conjuntos de datos obtenidos en la imputación múltiple.

Participantes y métodos

En este segundo artículo han participado pacientes que acuden a su médico de Atención Primaria por lumbalgia, cervicalgia o dorsalgia en cinco Comunidades Autónomas españolas: Baleares, Asturias, Cataluña, Murcia y Madrid.

Según protocolos publicados (26,57,59,66), los médicos de atención primaria del Sistema Nacional de Salud derivaron pacientes con criterios de indicación de NRT a unidades especializadas para la realización de la neuroreflejo terapia. Estos criterios son: dolor dorsal, lumbar o cervical con una intensidad igual o superior a 3 en una escala visual analógica entre 0 y 10 (0 representa no dolor y 10 el peor dolor imaginable) y una duración del dolor igual o superior a 14 días que no haya remitido con medicación ni otros tratamientos convencionales.

No se recomienda la realización de la NRT en pacientes con las siguientes características:

- Aquellos en los que el dolor en la zona cervical o lumbar se debe a golpes directos (como caídas o accidentes), fracturas, aplastamientos vertebrales o enfermedades “sistémicas” (es decir, no originadas en la espalda), como infección, cáncer o enfermedades reumáticas inflamatorias (como espondilitis anquilosante o reumatismo psoriásico).
- Que estén en uno de estos cuatro supuestos, en los que la cirugía de columna vertebral puede estar indicada:
 - “Síndrome de la cola de caballo” (pérdida muy importante o rápidamente progresiva de la fuerza en la pierna o piernas, pérdida de control de los esfínteres, o “anestesia en silla de montar” –pérdida de sensibilidad en la zona genital-).
 - Claudicación neurógena por estenosis espinal: corresponde al dolor progresivo en la pierna o piernas, que aparece exclusivamente al andar y cede rápidamente al estar sentado (o flexionado hacia adelante).

- Síndromes de “Brown Sequard” o de “mielopatía cervical”, en los que la médula es comprimida de manera uni o bilateral a nivel cervical, causando pérdida de sensibilidad o fuerza en brazos o piernas (y que pueden y suelen no causar dolor).
- Dolor irradiado (al brazo o la pierna), haya o no dolor cervical o lumbar, que es debido a la compresión de una raíz nerviosa por una hernia discal y dura 6 o más semanas pese a haber aplicado todos los tratamientos conservadores comprobadamente eficaces.

Pacientes con alguna de estas características no se incluyeron en el estudio. Los pacientes incluidos en este estudio son aquellos que fueron derivados con criterios de realización de la NRT y fueron dados de alta antes del 30 de junio del año 2012. Todos los pacientes que aceptaron participar en este estudio firmaron un consentimiento informado previamente.

El diseño del estudio correspondió a un estudio de cohortes prospectivo, con mediciones en la primera visita (visita basal) y a las 12 semanas, coincidiendo con la fecha de alta aproximada tras la realización de la intervención NRT. Los pacientes que, tras estas 12 semanas, cumplieron criterios para repetir la intervención NRT, fueron excluidos de este estudio, puesto que el objetivo era evaluar el pronóstico tras una única intervención.

Los datos del paciente, previos a la intervención de la NRT, fueron recogidos por los correspondientes médicos de Atención Primaria. Se recogió información sobre la edad y género de los pacientes, situación laboral (dividida en tres categorías: sin trabajo remunerado, en caso de ser estudiantes, amas de casa, voluntarios sin otros trabajos remunerados, jubilados o parados; de baja o invalidez por el dolor motivo de consulta; y trabajando de manera remunerada) y si se encontraba en algún proceso de reclamación laboral (por ejemplo, tramitación de la invalidez) o legal (por ejemplo, un accidente de tráfico) por el episodio de consulta.

Las variables que recabaron información clínica de los pacientes son: localización del dolor por el que consulta (lumbar, cervical o dorsal), tiempo transcurrido desde que

apareció el dolor por primera vez y desde que apareció el episodio por el que consulta, existencia de dolor irradiado, si el tipo de dolor es debido a compresión nerviosa por estenosis espinal o hernia/protrusión discal (definido como dolor referido más intenso que el dolor por el que se consulta y con hallazgo radiológico de estenosis espinal, hernia o protrusión discal), si ha sido intervenido quirúrgicamente del segmento por el que consulta, si ha sido diagnosticado de “síndrome de cirugía raquídea fallida” (ha sido intervenido quirúrgicamente y el dolor sigue igual o ha empeorado), diagnóstico de fibromialgia y otras comorbilidades (hipertensión, diabetes, insuficiencia hepática, renal, etc.), pruebas diagnósticas prescritas por el motivo de consulta (radiografía simple, resonancia magnética y otras pruebas como electromiograma o analítica sanguínea), hallazgos radiológicos en las pruebas de imagen (degeneración discal o artrosis, artrosis o hipertrofia de las articulaciones facetarias, escoliosis, disimetría de miembros inferiores, espondilólisis, espondilolistesis, estenosis espinal, protrusión discal o “hernia no extruida”, hernia discal o “hernia extruida” y otros hallazgos radiológicos), tratamientos farmacológicos prescritos para el episodio actual (analgésicos, opiáceos, antiinflamatorios no esteroideos, corticoides, relajantes musculares, codeína y fármacos de otro tipo), y otros tratamientos como la fisioterapia o la rehabilitación.

También se recogió información sobre la intensidad del dolor local (dolor por el que acudía a la consulta), dolor irradiado y nivel de discapacidad producida por el dolor, basalmente y a las 12 semanas.

Todas las escalas utilizadas para evaluar estas variables son escalas validadas y adaptadas al castellano. La intensidad del dolor local e irradiado se evalúa gracias a una escala visual-analógica (VAS) que oscila entre 0 y 10, representando 0 la ausencia de dolor y 10 el peor dolor imaginable (67). La discapacidad asociada a dolor lumbar se evalúa por medio del cuestionario Roland Morris, que mide el grado en el que el dolor provoca limitaciones en actividades cotidianas como vestirse, asearse o conducir, y que puntúa en un intervalo entre 0 y 24 (cuanta mayor puntuación, mayor nivel de discapacidad) (68). La discapacidad asociada al dolor cervical se evalúa con el Índice de Discapacidad Cervical (*Neck Disability Index*, NDI). Su intervalo oscila entre 0 y 100, constituyendo una puntuación mayor cuanto mayor es el grado de discapacidad. Este

cuestionario se validó y adaptó al castellano en el año 2008 por la REIDE, con la participación de la doctoranda en la fase de análisis (69)[§]. Hasta la fecha, no hay instrumentos que midan el nivel de discapacidad asociado al dolor dorsal, por lo que estos pacientes fueron excluidos del análisis de resultados.

El intervalo de medición para la discapacidad es diferente según el paciente haya acudido a la consulta por lumbalgia (oscila entre 0 y 24 puntos según el cuestionario Roland Morris), o cervicalgia (oscila entre 0 y 100 puntos según la escala NDI). Para poder manejar la variable “discapacidad” como una sola, se ha re-escalado la puntuación de discapacidad para lumbalgia de manera que oscile con un intervalo entre 0 y 100.

El nivel académico comenzó a recogerse a partir del año 2009. Se decidió excluir esta variable de los análisis, a pesar de estar recogida de forma parcial, porque la información faltante era demasiado abundante.

Posteriormente, ciertas variables se recodificaron para facilitar la interpretación de sus coeficientes en los modelos de regresión:

- Tiempo transcurrido desde que apareció el dolor por primera vez, se recodificó en cuatro categorías: igual o menor a un año, entre 1 y 5 años, entre 5 y 10 años y más de 10 años.
- Duración del presente episodio de dolor (en días): sub-agudo si es igual o menor a 90 días, crónico cuando oscila entre 91 y un año, y muy crónico si es superior a un año.

Estos datos se muestran en la Tabla 2.

[§]En el Apéndice IV se adjunta la correspondiente publicación.

Tabla 2. Variables recogidas en la visita basal

Características demográficas de los pacientes
-Género -Edad (años) -Situación laboral: Sin trabajo remunerado (estudiante, ama de casa, jubilado, parado) De baja o invalidez por el dolor motivo de consulta Trabajando de manera remunerada -Inmerso en proceso de reclamación laboral relacionado con este episodio (sí/no) -Inmerso en proceso de reclamación legal relacionado con este episodio (sí/no)
Variables relacionadas con el dolor lumbar o cervical
-Localización del dolor por el que está consultando (lumbar/cervical) -Tipo de dolor (dolor debido a compresión nerviosa por estenosis espinal o hernia-protrusión discal/ no) -Duración del episodio actual Sub-agudo (≤ 90 días) Crónico (91-365 días) Muy crónico (> 365 días) -Intensidad del dolor local (lumbar o cervical) (escala VAS: 0-10) -Intensidad dolor irradiado (escala VAS: 0-10) -Grado de discapacidad asociado al dolor lumbar o cervical (0-100)
Características clínicas de los pacientes y antecedentes
-Diagnóstico de fibromialgia (sí/no) -Diagnóstico de otras enfermedades (hipertensión, diabetes, etc.) -Duración del dolor desde la primera vez que apareció ≤ 1 año 1-5 años 5-10 años > 10 años -Cirugía previa del segmento doloroso (sí/no) -Síndrome de cirugía raquídea fallida (sí/no) -Procedimientos diagnósticos prescritos para este episodio: Rayos-X(sí/no) Resonancia Magnética(sí/no) Otros (electromiografía, analítica sangre)(sí/no) -Hallazgos radiológicos por este episodio Deshidratación o degeneración discal, o artrosis(sí/no) Degeneración o artrosis de las articulaciones facetarias (o unciformes)(sí/no) Escoliosis(sí/no) Dismetría miembros inferiores(sí/no) Espondilólisis(sí/no) Espondilolistesis(sí/no) Estenosis espinal(sí/no) Protrusión discal (o hernia no extruida)(sí/no) Hernia discal (extruida) (sí/no) Otros hallazgos radiológicos relevantes para este episodio(sí/no) Otros hallazgos radiológicos no relevantes para este episodio (Síndrome inespecífico)(sí/no) Ningún hallazgo radiológico(sí/no)
Tratamientos prescritos por este episodio
-Farmacológicos Analgésicos(sí/no) Antiinflamatorios no esteroideos (AINEs)(sí/no)

Corticoides(sí/no)
Relajantes musculares(sí/no)
Opiáceos(sí/no)
Codeína(sí/no)
Otros fármacos(sí/no)
-No farmacológicos
Fisioterapia o rehabilitación (ejercicio, técnicas de calor y/o frío, electroterapia, técnicas manuales como masaje o movilización)(sí/no)

La definición de los *outcomes* atiende al mismo criterio empleado en el primer artículo de esta Tesis. Se estudiaron tres variables resultado: “Mejoría en el dolor local”, “Mejoría en el dolor irradiado” y “Mejoría en la discapacidad”. En estas tres definiciones, se utilizó como punto de corte el correspondiente al cambio mínimo clínicamente relevante que, en términos generales, corresponde al 30% de la puntuación basal, con un valor mínimo de 1,5 puntos en la escala VAS, 2,5 puntos en la escala de Roland Morris y 7 puntos en la escala NDI (37,45,60). Con estas definiciones, en aquellos pacientes que tuvieran una puntuación basal, ya sea en intensidad de dolor o en discapacidad, inferior a estos puntos de corte, no era posible observar una mejoría clínicamente importante en ellos. Por tanto, se excluyeron del análisis para el correspondiente modelo. A modo de ejemplo, si un paciente puntuó su intensidad de dolor irradiado en 1, ese paciente fue excluido del modelo correspondiente a la mejoría del dolor irradiado, pero sí pudo participar en los modelos de mejoría del dolor local y discapacidad, siempre y cuando sus puntuaciones basales en intensidad de dolor local y discapacidad se encontraran por encima del punto de corte correspondiente.

Se desarrollaron tres modelos de regresión logística múltiple para cuantificar la capacidad predictiva de las variables recogidas en la visita basal. Estas variables fueron seleccionadas por su relevancia clínica y por su demostrada capacidad pronóstica en la literatura (44,54,57,59,67,70–73). Se incluyeron todas las variables recogidas en la visita basal salvo aquellas con una prevalencia inferior al 1%, para evitar inestabilidad numérica.

La imputación múltiple de datos faltantes se llevó a cabo por el método de las ecuaciones encadenadas, de la misma manera que en el primer artículo de esta Tesis. De nuevo, se asumió un patrón de pérdida de datos al azar (MAR), y en el proceso de

imputación se añadieron todas las variables basales que, bajo esta asunción, pudieran recoger la información necesaria para explicar la pérdida de datos. El número de archivos de datos imputados se fijó en 10 ($m=10$) para cada uno de los modelos pronósticos planteados, y los resultados correspondientes a cada una de las imputaciones múltiples se promediaron mediante las reglas de Rubin ^{**} (12).

Para la ejecución de este proceso se utilizaron los comandos *ice* y *mim* de Stata (17). Aquellas variables que tuvieran más de un 50% de datos faltantes, se excluyeron del proceso, ya que la imputación múltiple deja de ser apropiada (9).

Una vez que se obtuvieron los tres modelos finales, se dibujaron sus correspondientes nomogramas para facilitar su interpretación y mejorar su aplicabilidad. El paquete estadístico utilizado para este fin fue el software R y sus librerías *rms*, *car*, *survival* y *Hmisc* (48).

Se evaluó la validez interna de cada modelo mediante su calibración y discriminación correspondientes.

La calibración se valoró creando un gráfico de calibración en el que se representaron, para cada uno de los modelos desarrollados, la “Curva perfecta” (calibración perfecta), la “Curva aparente” (calibración de los datos originales), y la “Curva con el sesgo corregido” (calibración del modelo corregida por *bootstrapping*).

La técnica de *bootstrapping* permite estimar el optimismo del modelo, asumiendo que el modelo final no necesita calibración, es decir, en la curva de calibración, los valores de la ordenada en el origen y la pendiente son iguales a 0 y 1, respectivamente ($\gamma = (\gamma_0, \gamma_1) = (0, 1)$). Por tanto, calculando γ se consigue una estimación del sobre-optimismo.

Como en este trabajo se obtuvieron 10 conjuntos de datos derivados de la imputación múltiple en cada uno de los modelos, partimos de 10 valores estimados de $\gamma = (\gamma_0, \gamma_1)$. Por tanto, para conseguir una única “Curva aparente”, se promediaron los 10 valores estimados en cada caso. La “Curva con el sesgo corregido” se obtuvo mediante 500 muestreos *bootstrap*. Todas las curvas se dibujaron gracias al comando *lowess* de Stata, que se encarga de proporcionar medias suavizadas. El gráfico situado en la parte

^{**} En el Apéndice II se encuentra una especificación detallada de las reglas de Rubin.

inferior (*rugplot*) indica la distribución de probabilidades predichas del correspondiente modelo.

Como medida de la discriminación se utilizó el *índice D_{xy} de Somer*, que se relaciona con el *índice c* de concordancia gracias a la fórmula $D_{xy} = 2(c - 0,5)$ (6). El *índice c* estima la probabilidad de concordancia entre respuestas observadas y predichas. Al representar lo mismo que el área bajo la curva ROC, su interpretación es la misma. Por tanto, un valor igual a 0,5 indica nula capacidad discriminatoria, y un valor igual a 1 indica una perfecta separación entre aquellos pacientes que presentan el evento de interés y los que no. El *índice D_{xy} de Somer*, supone un cambio de escala hacia un índice de correlación de rangos entre las probabilidades predichas y las respuestas observadas, con un intervalo entre -1 y +1, y un valor igual a 0 cuando no hay correlación. El *índice D_{xy}* debe ser interpretado con cautela, ya que el modelo ajustado probablemente será sobre-optimista. Teniendo en cuenta ese posible sobre-optimismo, el *índice D_{xy} de Somer* se estimó por medio de 500 réplicas *bootstrap*.

Resultados

Los resultados presentados a continuación corresponden al artículo *“Predicting outcomes of neuroreflexotherapy in patients with subacute or chronic neck or low back pain”*, aceptado el 19 de septiembre de 2013 en la revista *The Spine Journal* y disponible *online* a partir del 18 de octubre del mismo año. El día en que se deposita esta Tesis Doctoral, todavía no se dispone de la versión en papel del artículo.

Para facilitar la interpretación de los coeficientes de los tres modelos pronósticos, se incluyeron en la publicación los tres nomogramas correspondientes. Atendiendo a las características de cada paciente, sólo hay que sumar los puntos que correspondan, tal y como se indica en las figuras 2, 3 y 4 del artículo publicado (páginas 8-10).

Hace unos años, Lenfant(50) escribía un artículo especial en el que analizaba el valle existente entre la investigación y la práctica clínica, y citaba textualmente: “Tanto los investigadores médicos como los líderes políticos hablan sobre la falta de éxito a la hora de trasladar los hallazgos científicos a la práctica médica y el comportamiento personal. Independientemente de las razones para este fenómeno, estructurales, económicas o de motivación, el resultado sigue siendo el mismo: no estamos recogiendo todos los beneficios en cuanto a salud pública de nuestra inversión en investigación”.

A partir de este gran reto, esta Tesis intenta alcanzar el objetivo de acercar los descubrimientos de la investigación clínica a la práctica médica. A partir de este segundo artículo, también se ha desarrollado una aplicación web que facilite la accesibilidad e interpretación de los resultados de los modelos pronósticos.

Puesto que el artículo se encuentra todavía “en prensa” el día en que se deposita esta Tesis, la aplicación web no está disponible para el público. En el momento en que salga publicado el artículo, la herramienta web se subirá al servidor de la Fundación Kovacs y divulgará mediante una rueda de prensa en la Organización Médica Colegial (OMC) en Madrid.

Se trata de una herramienta accesible tanto a pacientes como a profesionales, que estima la probabilidad de mejora clínicamente relevante en cuanto a la intensidad de dolor cervical o lumbar, intensidad de dolor irradiado (al brazo, en el caso del dolor cervical, o a la pierna -“ciática”- en el caso del dolor lumbar), y discapacidad, definida como grado de restricción de las actividades cotidianas que impone el dolor, tres meses después de haber sido intervenidos con la NRT.

El paciente o el médico, sólo tiene que introducir los valores de las variables solicitadas en la aplicación (Figura3), seleccionar el botón de “Calcular” y la aplicación mostrará el resultado en cada caso (Figura 4).

La aplicación cuenta con una ayuda que sirve de tutorial para conocer a quién va dirigida esta calculadora y a quien no, en qué está basado su cálculo, citando el presente trabajo de investigación, así como sus limitaciones. También cuenta con una aclaración a cada una de las preguntas que deben completarse para obtener el resultado final.

A modo de ejemplo, un profesor de instituto con 40 años de edad, que acude a la consulta con dolor lumbar desde aproximadamente mes y medio, fija su intensidad de dolor lumbar en 7 en una escala visual analógica (VAS), una intensidad de dolor irradiado igual a 3 y un grado de discapacidad igual a 15 puntos en la escala Roland Morris. Explica que le realizaron una radiografía simple por ese mismo dolor hace quince días y le diagnosticaron una escoliosis. Ha estado tomando analgésicos y antiinflamatorios para calmar el dolor desde que apareció. Tras haberle realizado la intervención NRT, este paciente tiene una probabilidad cercana al 88% en cuanto a mejorar en su dolor lumbar de forma clínicamente relevante dentro de tres meses, un 87% en cuanto a mejorar el dolor irradiado a la pierna y un 62% para mejorar su discapacidad (Figura 4).

Figura 3. Visualización de la aplicación web a la hora de introducir la información sobre las diferentes variables.

<p>Edad ⓘ</p> <div style="border: 1px solid #ccc; padding: 2px; width: 100px; text-align: center;">40</div> <p>Sexo ⓘ</p> <p><input checked="" type="radio"/> Masculino <input type="radio"/> Femenino</p> <p>Localización del dolor por el que está consultando ⓘ</p> <p><input type="radio"/> Cervical <input checked="" type="radio"/> Lumbar</p> <p>Tipo de dolor ⓘ</p> <p><input type="radio"/> Dolor debido a compresión nerviosa por estenosis espinal o hernia/protrusión discal <input checked="" type="radio"/> Otro tipo de dolor</p> <p>Situación laboral ⓘ</p> <p><input type="radio"/> Pasiva <input type="radio"/> De baja o invalidez específicamente por el dolor cuyo pronóstico está calculando <input checked="" type="radio"/> Activa (trabajando)</p> <p>Tiempo transcurrido desde que apareció el dolor por primera vez (en años) ⓘ</p> <p><input checked="" type="radio"/> Un año o menos <input type="radio"/> Más de un año pero menos de 5 <input type="radio"/> Entre 5 y 10 años <input type="radio"/> Más de 10 años</p>	<p>Duración del presente episodio de dolor (en días) ⓘ</p> <p><input checked="" type="radio"/> 90 días o menos <input type="radio"/> Entre 91 y 365 días <input type="radio"/> Más de 365 días</p> <p>¿Padece (además) fibromialgia? ⓘ</p> <p><input type="radio"/> Sí <input checked="" type="radio"/> No</p> <p>¿Padece (además) otras enfermedades? ⓘ</p> <p><input type="radio"/> Sí <input checked="" type="radio"/> No</p> <p>¿Está en proceso de reclamación laboral por el dolor cuya probabilidad de mejorar tras la intervención NRT está calculando? ⓘ</p> <p><input type="radio"/> Sí <input checked="" type="radio"/> No</p> <p>¿Ha sido intervenido quirúrgicamente previamente del segmento que actualmente le duele? ⓘ</p> <p><input type="radio"/> Sí <input checked="" type="radio"/> No</p> <p>¿Le han diagnosticado que su dolor se debe a “síndrome de cirugía raquídea fallida”, “fibrosis post-quirúrgica” o “síndrome post-laminectomía”? ⓘ</p> <p><input type="radio"/> Sí <input checked="" type="radio"/> No</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 3(continuación). Visualización de la aplicación web a la hora de introducir la información sobre las diferentes variables.

¿Le han prescrito una radiografía (simple) para diagnosticar la causa del episodio doloroso que está padeciendo?

☒ Sí
☐ No

¿Le han prescrito una resonancia magnética para diagnosticar la causa del episodio doloroso que está padeciendo?

☐ Sí
☒ No

¿Le han prescrito alguna otra prueba diagnóstica por el episodio doloroso que está padeciendo (electromiograma, analítica sanguínea, etc.)?

☐ Sí
☒ No

Resultados de las pruebas de imagen, si se han realizado

☐ Deshidratación o degeneración discal, o artrosis
☐ Degeneración, artrosis o hipertrofia de las articulaciones facetarias (o unciformes)
☒ Escoliosis
☐ Espondilólisis
☐ Espondilolistesis
☐ Estenosis espinal
☐ Protrusión discal (o "hernia no extruida")
☐ Hernia discal (o "hernia extruida")
☐ Otros hallazgos radiológicos que **sí** puedan ser relevantes para su dolencia actual
☐ Otros hallazgos radiológicos que **no** son relevantes para su dolencia actual
☐ Ningún hallazgo radiológico

Tratamientos prescritos por el episodio doloroso

☒ Analgésicos (como paracetamol)
☐ Analgésicos opiáceos (como codeína, tramadol, etc.)
☒ Antiinflamatorios no esteroideos (como ibuprofeno, diclofenaco, etc.)
☐ Corticoides
☐ Relajantes musculares
☐ Fármacos de otro tipo
☐ Rehabilitación o fisioterapia

¿De qué intensidad es el dolor que está sintiendo hoy en el segmento por el que está consultando (cervical o lumbar)?

Ausencia de dolor Máximo dolor imaginable

7

¿De qué intensidad es el dolor irradiado (al brazo –en caso de dolencias cervicales- o a la pierna –en caso de dolencias lumbares-) que está sintiendo hoy?

Ausencia de dolor Máximo dolor imaginable

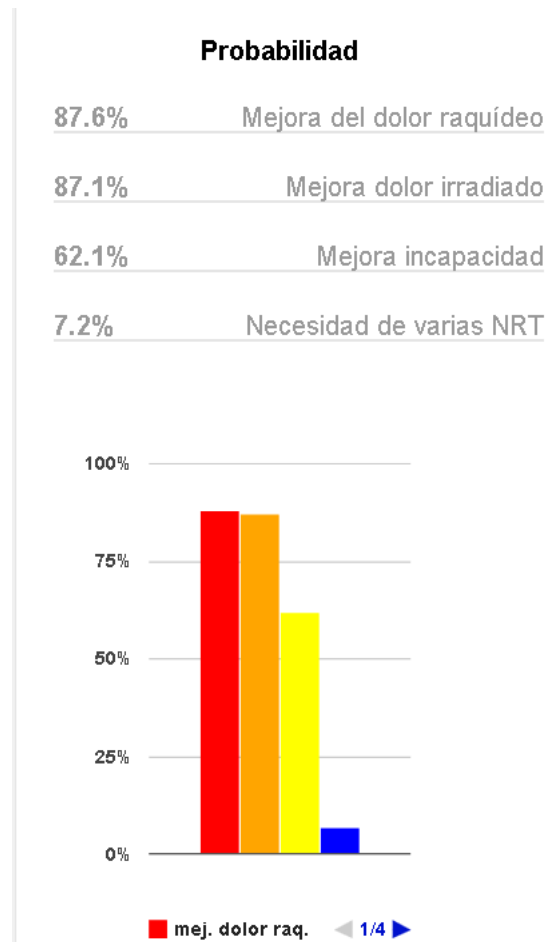
3

¿Cuál es el grado de incapacidad (o "restricción de las actividades cotidianas") que causa el dolor lumbar? [RMQ]

15

Calcular

Figura 4. Visualización de la aplicación web mostrando la probabilidad de mejoría para cada uno de los tres modelos en un caso concreto.



La generación del gráfico de calibración cuyo desarrollo tiene en cuenta los diferentes conjuntos de datos derivados del proceso de imputación múltiple, dio lugar a la presentación de un póster en el Congreso Europeo de Epidemiología, EUROEPI 2013, celebrado en la ciudad danesa de Aarhus, en Agosto de 2013. El título del poster es *“Calibration Plot Joining Multiple Imputation Datasets”*.

En el Apéndice V de esta Tesis Doctoral se adjunta una copia del poster presentado.



Clinical Study

Predicting outcomes of neuroreflexotherapy in patients with subacute or chronic neck or low back pain

Ana Royuela, MSc^{a,b,c,*}, Francisco M. Kovacs, MD, PhD^{c,d}, Carlos Campillo, MD, MPH, PhD^e,
Montserrat Casamitjana, MD^f, Alfonso Muriel, MSc^{b,c}, Víctor Abaira, PhD^{a,b,c}

^aCIBER Epidemiología y Salud Pública (CIBERESP), C/ Melchor Fernandez Almagro 3-5, 28029 Madrid, Spain

^bUnidad de Bioestadística Clínica, Hospital Ramón y Cajal, IRICYS, C/ Colmenar Viejo, 9, 28031 Madrid, Spain

^cSpanish Back Pain Research Network, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain

^dDepartamento Científico, Fundación Kovacs, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain

^eServei de Salut de les Illes Balears (Ib-Salut), Calle Reina Esclaramunda 9, 07003 Palma de Majorca, Spain

^fRegió Sanitària de Barcelona | Consorci Sanitari de Barcelona, Servei Català de la Salut (CatSalut), Parc Sanitari Pere Virgili - Edifici Mestral - Esteve Terradas, 30 4ta planta, 08023 Barcelona, Spain

Received 1 October 2012; revised 27 August 2013; accepted 19 September 2013

Abstract

BACKGROUND CONTEXT: In the context of shared decision-making, a valid estimation of the probability that a given patient will improve after a specific treatment is valuable.

PURPOSE: To develop models that predict the improvement of spinal pain, referred pain, and disability in patients with subacute or chronic neck or low back pain undergoing a conservative treatment.

STUDY DESIGN AND SETTING: Analysis of data from a prospective registry in routine practice.

PATIENT SAMPLE: All patients who had been discharged after receiving a conservative treatment within the Spanish National Health Service (SNHS) (n=8,778).

OUTCOME MEASURES: Spinal pain, referred pain, and disability were assessed before the conservative treatment and at discharge by the use of previously validated methods.

METHODS: Improvement in spinal pain, referred pain, and disability was defined as a reduction in score greater than the minimal clinically important change. A predictive model that included demographic, clinical, and work-related variables was developed for each outcome using multivariate logistic regression. Missing data were addressed using multiple imputation. Discrimination and calibration were assessed for each model. The models were validated by bootstrap, and nomograms were developed.

RESULTS: The following variables showed a predictive value in the three models: baseline scores for pain and disability, pain duration, having undergone X-ray, having undergone spine surgery, and receiving financial assistance for neck or low back pain. Discrimination of the three models ranged from slight to moderate, and calibration was good.

CONCLUSIONS: A registry in routine practice can be used to develop models that estimate the probability of improvement for each individual patient undergoing a specific form of treatment. Generalizing this approach to other treatments can be valuable for shared decision making. © 2013 Elsevier Inc. All rights reserved.

Keywords:

Predictive model; Multiple imputation; Calibration; Back pain; Disability; Neuroreflexotherapy

FDA device/drug status: Not applicable.

Author disclosures: **AR:** Nothing to disclose. **FMK:** Nothing to disclose. **CC:** Nothing to disclose. **MC:** Nothing to disclose. **AM:** Nothing to disclose. **VA:** Nothing to disclose.

This study was funded by the Kovacs Foundation, a Spanish not for profit Institution with no links to the health industry. The funding institutions had no role in the design and conduction of the study; data collection;

management, analysis and interpretation of the data; preparation, review and approval of the manuscript; or the decision to submit the article for publication. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this article.

* Corresponding author. Clinical Biostatistics Unit, Hospital Ramón y Cajal, Ctra. Colmenar Km 9.1, 28034 Madrid, Spain. Tel.: (34) 913368103.
E-mail address: ana.royuela@hrc.es (A. Royuela)

Introduction

In industrialized countries, neck and common low back pain (LBP) affect more than 70% of the general population and represent a major health, social, and economic burden [1–8]. Diagnosing common neck pain or LBP implies that the pain is not related to fractures, direct trauma, or systemic conditions such as spondylitis or neoplastic-, infectious-, vascular-, metabolic-, visceral-, or endocrine-related processes. In more than 90% of the patients, it is impossible to identify the exact cause of pain, which is believed to originate in soft tissues [1,9–13]. In routine practice, many treatments are prescribed for patients with neck pain and LBP, although very few are based on solid evidence on efficacy, effectiveness, safety, and cost/effectiveness [9,11,14].

Neuroreflexotherapy (NRT) is an intervention consisting in the temporary (up to 90 days), subcutaneous implantation of surgical devices on trigger points in the back, at the site of dermatomes clinically involved in each case, and on referred tender points located in the ear [15–22]. The intervention is assumed to deactivate neurons involved in the mechanisms that prolong pain, neurogenic inflammation, and muscle dysfunction and contracture [15–18]. The evidence available, including a Cochrane review, shows that it is efficacious, safe, effective, and cost-effective for improving pain and disability in subacute and chronic LBP patients [9,15–18]. A registry that uses test-piloted postimplantation surveillance methods was set up in 2002, when NRT was implemented in the Spanish National Health Service (SNHS) [19–22].

In clinical practice, there is always some uncertainty as to how results from randomized controlled trials will apply to an individual patient. In the context of sharing decision-making, it would be suitable to reduce this uncertainty by informing each patient of the probability that each treatment will trigger a clinically relevant improvement in his/her specific case. This would allow patients to balance this probability against personal preferences and the risk and gravity of adverse events. To this end, it would be valuable to explore the feasibility of using registries in routine practice to develop valid predictive models for the results of each form of treatment.

Registries and other observational studies in routine practice frequently present a high proportion of missing data, which can affect the validity of results. Multiple imputation by chained equations has been considered an appropriate method to address this problem [23–26]. A predictive model requires validation before considering its use in routine clinical practice [27,28]; bootstrapping is considered to be the most efficient validation procedure [29,30].

Therefore, the objective of this study was to assess the feasibility of using a registry in routine practice to develop models predicting the probability of improvement following a specific form of conservative treatment, using

multiple imputation to manage missing data and bootstrapping for validating the models.

Material and methods

Setting

This study was conducted in the SNHS of the Balearic Islands (“Ib-Salut”), Asturias (“SESPA”), Cataluña (“Cat-Salut”), Murcia (“SMS”), and Madrid (“SERMAS”). The population covered by these Health Services is approximately 17.6 million inhabitants, or approximately 37.5% of the Spanish population [31].

Study population

Following the referral protocol which was used in previous studies and current evidence-based Clinical Guidelines [9,19–21], primary care physicians referred patients with indication criteria for NRT to NRT-certified units. Indication criteria were as follows: common neck, thoracic or LBP, pain severity ≥ 3 points on a 10-point visual analog scale (VAS) [32], and pain lasting ≥ 14 days despite medication and other treatment. Referral should not be considered for patients with neurogenic claudication caused by lumbar spinal stenosis or those showing criteria for urgent referral to surgery, such as signs suggesting cauda equina syndrome (eg, progressive motor weakness in the legs, sphincter disturbance, saddle anesthesia, or sensory level). (A flowchart containing the referral protocol to NRT is available online [1].) Patients who did not respond to back surgery could be referred to NRT, as well as those with “red flags” in whom the appropriate tests had established that pain was not caused by fractures or systemic diseases [1]. All SNHS patients treated with NRT who were discharged before June 30, 2012, were included in this study.

Intervention

NRT interventions were performed according to the indication criteria and under the application conditions in which it had proven effective, safe, and cost-effective [15,19,20]. Specialists at the NRT units confirmed indication criteria and, once patients had given their written informed consent, performed the interventions. Each SNHS region had its own NRT unit and team of physicians, certified according to specialized professional standards [33]. Physicians had been in full-time NRT practice for between 3 and 23 years.

Patients were instructed to contact the NRT unit or their primary care center in the event of experiencing an adverse effect. Twelve weeks after the intervention was performed, the surgical material that had been implanted was removed, adverse events identified by patients or physicians were recorded, and indication criteria for repeating NRT were assessed.

Outcome assessment

Standardized postimplantation surveillance mechanisms were used to gather the data introduced into the registry [19,20]. The clinical condition of each patient upon referral to NRT was assessed by the referring primary care physician, who gathered data on duration of pain (days since the first episode and for the current episode, separately); reason for referral (neck pain, thoracic pain or LBP); existence of referred pain (yes/no); history of spine surgery; diagnosis of failed surgery syndrome; pain caused by symptomatic disc protrusion or herniation (defined as referred pain being more intense than spinal pain and after a distribution that corresponded to the root compressed by a disc protrusion/herniation on magnetic resonance imaging [MRI]); pain caused by symptomatic lumbar spinal stenosis (defined as referred pain corresponding to the root/s compressed by lumbar spinal stenosis on MRI) (yes/no); “common” syndrome (defined as no clinical signs of nerve compression on physical examination and MRI); established diagnosis of fibromyalgia; comorbidities; diagnostic tests undertaken (X-rays, MRI, other, eg, electromyogram or scintigraphy); imaging findings (disc degeneration, facet joint degeneration, scoliosis, difference in leg length, spondylolysis, spondylolisthesis, spinal stenosis, disc protrusion, disc herniation, other findings, no findings); treatments received for the current episode (drugs [eg, analgesics, nonsteroidal anti-inflammatory drugs [NSAIDs], steroids, muscle relaxants, opioids, codeine, other], physiotherapy/rehabilitation, surgery); and date of referral to the NRT unit.

Patients provided data on gender, age (date of birth), whether they were pregnant, employment status (“passive,” eg, housewife, student, or retired; “receiving financial assistance for neck pain or LBP,” eg, sick leave or workers compensation benefits; or “working”), involvement in neck pain or LBP-related employment claims (eg, disability pension), and involvement in neck pain or LBP-related litigation (eg, traffic accident). From May 2009 onwards, they were also requested to provide data on their academic level (less than elementary school, elementary school, high school, university).

Primary outcomes were pain and disability [34]. Patients used previously validated instruments to score these variables. Separate 10-cm VAS were used for spinal pain (neck pain or LBP) and referred pain [32]. The Spanish version of the Roland-Morris Questionnaire (RMQ) was used to score LBP-related disability [36]. The Neck Disability Index (NDI) was used to assess neck pain-related disability in patients treated after its Spanish version was validated, in April 2008 [36]. Value ranges from best to worst are 0–10 for VAS, 0–24 for RMQ, and 0–100 for NDI [32,35,36].

The severity of spinal pain, referred pain, and disability were assessed on each visit to the primary care center and specialized NRT unit. All diagnostic and therapeutic procedures undergone after NRT recorded throughout the follow-up period. Twelve weeks after performing the

NRT intervention, pain and disability were assessed, and patients without indication criteria for repeating the procedure were discharged.

Analysis

Absolute and relative frequencies were calculated for categorical variables. Values for continuous variables were described using their median and interquartile range. It was hypothesized that different pain episodes in the same subject (either at the same or different locations, eg, neck pain and LBP) could be nonindependent events. Therefore, analyses were restricted to the first episode treated with NRT in each patient.

“Improvement” in spinal pain, referred pain, and disability was defined as any reduction in the corresponding score being greater than the minimal clinically important change (MCIC) [37–39]. MCICs have been established at 30% of the baseline value, with a minimum value of 1.5 VAS points for spinal pain and referred pain, 7 NDI points for neck pain-related disability, and 2.5 RMQ points for LBP-related disability [37–39]. These definitions made it impossible for patients with a baseline score less than the corresponding cut-off point for a given variable to show a clinically relevant improvement for that variable (eg, referred pain could not improve in patients without referred pain at baseline). Therefore, these patients were excluded from the analysis on that variable.

Three separate multivariate logistic regression models were developed to predict improvement of spinal pain, referred pain, and disability, respectively. Because value ranges are different for the NDI (0–100) and the RMQ (0–24), a standardized score for disability was calculated, ranging from 0 to 100 (from better to worse). For neck pain-related disability, this score corresponded to the NDI score. For LBP-related disability, it corresponded to the percentage of the maximum possible RMQ score (eg, 24 points in the RMQ corresponded to 100 points in the standardized score). Because no instrument for assessing disability derived from thoracic pain has been validated in Spanish, patients with thoracic pain were excluded from the models.

In these models, improvement in spinal pain, referred pain and disability were the dependent variables. Independent variables were those considered to be clinically relevant and those that had shown to have a prognostic value [17,19,20,32,34,40–42]: gender, age (in years), baseline score for spinal pain (VAS points), baseline score for referred pain (VAS points), baseline score for disability (standardized disability score), reason for referral (neck pain or LBP), time elapsed since the first pain episode (<1 year, 1 to <5 years, 5 to <10 years, ≥10 years), duration of the current episode (“subacute” if between 14 and 89 days, “chronic” if between 90 and 365 days, and highly chronic if more than 365 days) [22,43,44], employment status (“passive”, “receiving financial assistance for neck pain or LBP”, or “working”), type of pain (“radicular pain

Table 1

Baseline characteristics of patients included in the study (N=8,778)

Variables	n Filled	Value
Gender, male*	8,694	2,775 (31.9)
Age, years [†]	8,774	53 (42; 64)
Region within the Spanish National Health Service*	8,778	
Balearic Islands (Ib Salut)		7,033 (80.1)
Asturias (SESPA)		765 (8.7)
Murcia (SMS)		484 (5.5)
Madrid (SERMAS)		382 (4.4)
Catalonia (Cat Salut)		114 (1.3)
Reason for referral to NRT*	8,666	
Neck pain		2,588 (29.9)
Low back pain		6,078 (70.1)
Type of pain ^{*,‡}	8,552	
Nonspecific		8,097 (94.7)
Radicular pain caused by disc protrusion/extrusion or spinal stenosis		455 (5.3)
Employment status*	7,682	
Passive		3,103 (40.4)
Receiving financial assistance for low back pain		806 (10.5)
Working		3,773 (49.1)
Duration of the pain since diagnosis, months [†]	8,371	73 (24.3; 146)
Duration of the pain since diagnosis categorized, years*		
≤1		1,360 (16.2)
1–5		2,663 (31.8)
5.001–10		2,125 (25.4)
>10		2,223 (26.6)
Duration of the pain episode, days [†]	8,778	300 (90; 540)
Duration of the pain episode, days, categorized*		
Subacute, ≤90		2,2116 (25.2)
Chronic, 91–365		3,932 (44.8)
Highly chronic, >365		2,635 (30.0)
Pregnancy*	8,778	32 (0.4)
Diagnosis of fibromyalgia*	8,778	305 (3.5)
Other comorbidities*	8,778	3,579 (40.8)
Involved in work-related claims*	8,778	67 (0.8)
Involved in litigation*	8,778	35 (0.4)
Baseline severity of SP (VAS) [†]	8,674	7.0 (6.0; 8.0)
Baseline severity of RP (VAS) [†]	8,555	7.0 (5.0; 8.0)
Baseline disability (standardized 0–100 score) [†]	7,217	54.2 (37.5; 70.8)
Previous lumbar surgery* (yes)	8,778	676 (7.7)
Failed back syndrome* (yes)	8,765	102 (1.2)
Diagnostic procedures during the episode*		
X-ray	8,778	2,196 (25.0)
MRI	8,778	2,701 (30.8)
Other [§]	8,778	592 (6.7)

MRI, magnetic resonance imaging; NDI, Neck Disability Index; RMQ, Roland-Morris Questionnaire; RP, severity of referred pain (in the 6,649 patients who had it); SP, severity of spinal pain; VAS, visual analog scale (range from better to worse; 0–10).

Note: Baseline disability (in the 6,766 patients who had it; range from better to worse: 0–100; for neck pain–related disability, the score of the NDI was used. For low back pain–related disability, the score reflects the percentage of the maximum possible RMQ score).

* Frequency (%).

[†] Median (percentile 25; percentile 75).

[‡] Type of pain: “Radicular pain caused by disc protrusion/extrusion or spinal stenosis” if; (a) severity of referred pain ≥ spinal pain, (b) corresponding imaging finding on MRI, (c) distribution of pain consistent with the nerve root compressed by the corresponding imaging finding. “Nonspecific pain,” if one or more of these criteria were not met.

[§] Other diagnostic procedures: electromyography, computed tomography scan, scintigraphy, and other.

caused by symptomatic disc protrusion/herniation or lumbar spinal stenosis” vs. “common neck pain or LBP”), diagnosis of fibromyalgia, other comorbidities, involvement in employment claims, involvement in litigation, diagnostic tests undertaken before NRT (X-rays, MRI, other), imaging findings, history of spine surgery, and treatments before referral for NRT. To avoid number instability,

variables with a prevalence of <1% were eliminated from the models.

A multiple imputation analysis was performed by creating several plausible imputed datasets and combining results from each of them [23,26]. Imputation of missing values was carried out by means of multiple imputation using chained equations and assuming a missing at random pattern of

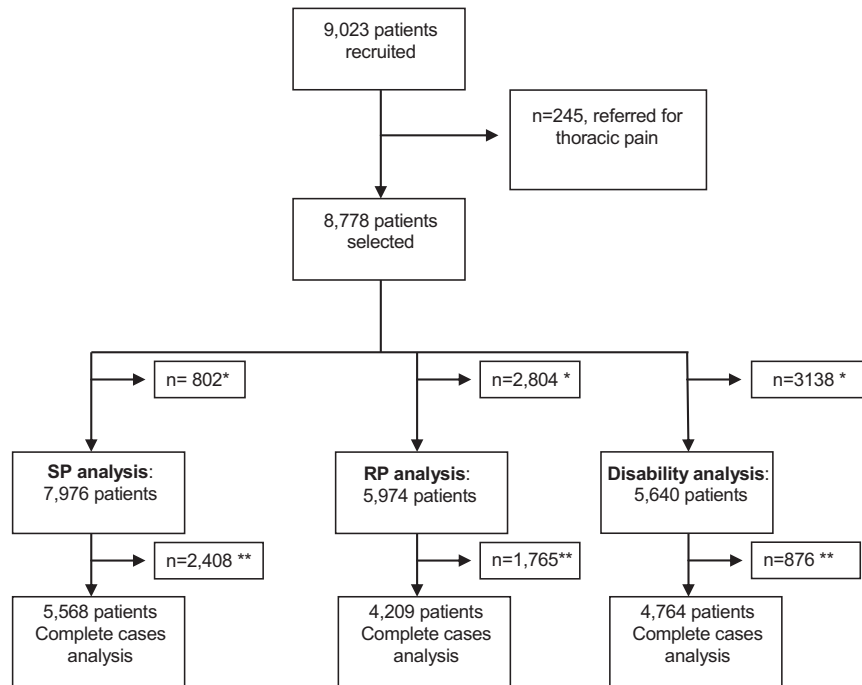


Fig. 1. Flowchart showing the number of patients analyzed. *Patients excluded because their baseline scores were below the cut-off points identifying potential improvement as “clinically relevant.” **Patients with at least one missing data in any of the variables included in the corresponding model.

missingness [23,24]. A total of 10 imputed datasets were created [45]. Rubin rules were used to average the regression coefficients and standard errors of the predictors in the final model [46]. *Ice* and *mim* Stata commands were used to perform the multiple imputation and analyze the imputed datasets, respectively [23]. At the design phase of the study, it was established that variables with $\geq 50\%$ missing data would be excluded from the analyses, because it would be inappropriate to use multiple imputation analysis [23].

To assess the accuracy of the final model, both their discrimination and calibration were evaluated. We assessed discrimination by estimating Somer's D rank correlation index, correcting for optimism by using bootstrap sampling (500 replicates). This index reflects the correlation between predicted probabilities and observed responses. Values range between -1 and $+1$. D_{xy} equals 0 when the model makes random predictions, and it equals 1 when predictions are perfectly discriminating. The *validate* function in the R's *rms* package was used [29,47]. Rubin rules were followed to combine the Somer's D indexes from each of the 10 imputed datasets.

To assess calibration, a calibration plot was developed [47], showing the “Apparent curve” (representing predicted probability using the model against actual probability), the “ideal curve” (representing perfect calibration, ie, predicted probability equals actual probability) [29], and the “bias-corrected curve” (representing the overfitting-corrected calibration curve, obtained by bootstrap resampling).

Bootstrapping makes it possible to estimate the overoptimism in assuming that the final model needs no calibration, that is, that it has overall intercept and slope

corrections of 0 and 1, respectively ($\gamma=(\gamma_0, \gamma_1)=(0,1)$). Therefore, γ was calculated to estimate overoptimism [29].

Because 10 imputed datasets had been created for each model (spinal pain, referred pain, and disability), 10 estimated values of $\gamma=(\gamma_0, \gamma_1)$ were averaged to calculate a global calibration curve in each case. Curves were drawn using the *lowess* function of Stata, which provides locally weighted scatterplot smoothing.

Nomograms were drawn to illustrate the results of these models [29]. Stata v11.0 (College Station, TX, USA) and R v2.15 software (available at: <http://www.R-project.org>) were used for the statistical analysis.

Results

Data from 9,023 patients were available. Two hundred forty-five patients (2.7%) had been referred for thoracic pain and were excluded. Therefore, 8,778 patients were included in this study. Most of these patients were women (68.1%) with chronic (75.0%) LBP (70%). The median baseline value for pain severity (for both spinal pain and referred pain) was 7.0 VAS points, and the median baseline standardized score for disability was 54.2% (Table 1).

Among the 8,778 patients included, 3,003 (34.3%) had been recruited after May 2009. Therefore, data on academic level were missing for 65.7% of the sample, so this variable was excluded from the analyses. Some patients also were excluded from their respective models on spinal pain, referred pain or disability, because their baseline scores were lower than the corresponding minimal clinically relevant change (Fig. 1).

Table 2
 Characteristics of patients who showed and did not show clinically relevant improvements in SP, RP, and disability after NRT

Variables	Spinal pain (n=5,568)		Referred pain (n=4,209)		Disability (n=4,764)	
	Shown a clinically relevant improvement (n=4,174)	Did not show a clinically relevant improvement (n=1,394)	Shown a clinically relevant improvement (n=2,996)	Did not show a clinically relevant improvement (n=1,213)	Shown a clinically relevant improvement (n=3,160)	Did not show a clinically relevant improvement (n=1,604)
Gender, male*	1,370 (32.8)	494 (35.4)	970 (32.4)	379 (31.2)	1,088 (34.4)	526 (32.8)
Age, years†	53 (42; 64)	52 (43; 64)	54 (43; 65)	53 (44; 64)	52 (42; 64)	54 (44; 64)
Reason for referral to NRT*						
Neck pain	742 (17.8)	230 (16.5)	529 (17.7)	196 (16.2)	344 (10.9)	234 (14.6)
Low back pain	3,432 (82.2)	1,164 (83.5)	2,467 (82.3)	1,017 (83.8)	2,816 (89.1)	1,370 (85.4)
Type of pain*‡						
Nonspecific	3,991 (95.6)	1,307 (93.8)	2,794 (93.3)	1,128 (93.0)	3,012 (95.3)	1,501 (93.6)
Radicular pain caused by disc protrusion/extrusion or spinal stenosis	183 (4.4)	87 (6.2)	202 (6.7)	85 (7.0)	148 (4.7)	103 (6.4)
Employment status*						
Passive	1,701 (40.8)	568 (40.7)	1,234 (41.2)	521 (43.0)	1,215 (38.4)	694 (43.3)
Receiving financial assistance for neck pain or low back pain	417 (10.0)	203 (14.6)	296 (9.9)	178 (14.7)	327 (10.3)	215 (13.4)
Working	2,056 (49.3)	623 (44.7)	1,466 (48.9)	514 (42.4)	1,618 (51.3)	695 (43.3)
Duration of the pain since diagnostic categorized, years*						
≤1	687 (16.5)	196 (14.1)	470 (15.7)	168 (13.8)	530 (16.8)	222 (13.8)
1–5	1,313 (31.5)	456 (32.7)	944 (31.5)	396 (32.6)	1,014 (32.1)	504 (31.4)
6–10	1,046 (25.1)	356 (25.5)	773 (25.8)	312 (25.7)	782 (24.7)	407 (25.4)
>10	1,128 (27.0)	386 (27.7)	809 (27.0)	337 (27.8)	834 (26.4)	471 (29.4)
Duration of the pain episode, days, categorized*						
Subacute, ≤90	1,210 (29.0)	266 (19.1)	833 (27.8)	216 (17.8)	947 (30.0)	344 (21.4)
Chronic, 91–365	1,818 (43.6)	674 (48.4)	1,312 (43.8)	593 (48.9)	1,349 (42.7)	763 (47.6)
Highly chronic, >365	1,146 (27.5)	454 (32.6)	851 (28.4)	404 (33.3)	864 (27.3)	497 (31.0)
Diagnosis of fibromyalgia*	135 (3.2)	50 (3.6)	99 (3.3)	51 (4.2)	90 (2.8)	61 (3.8)
Other comorbidities*	1,546 (37.0)	610 (43.8)	1,101 (36.7)	551 (45.4)	1,088 (34.4)	697 (43.5)
Involved in work-related claims*	18 (0.4)	20 (1.4)	18 (0.6)	15 (1.2)	17 (0.5)	17 (1.1)
Baseline severity of SP (VAS)†	7 (6; 8)	7 (6; 8)	7 (6; 9)	7 (6; 9)	7 (5; 8)	7 (6; 8)
Baseline severity of RP (VAS)†	7 (5; 8)	7 (5; 8)	7 (5; 8)	7 (5; 8)	7 (5; 8)	7 (5; 8)
Baseline disability (standardized 0–100 score)†	58.3 (41.7; 75.0)	52.0 (37.5; 70.8)	54.2 (38.0; 70.8)	60.0 (44.0; 75.0)	54.2 (37.5; 70.8)	54.2 (37.5; 70.8)
Previous lumbar surgery* (yes)	286 (6.9)	174 (12.5)	231 (7.7)	157 (12.9)	223 (7.1)	190 (11.8)
Failed back syndrome* (yes)	41 (1.0)	26 (1.9)	33 (1.1)	18 (1.5)	28 (0.9)	33 (2.1)
Diagnostic procedures during the episode*						
X-ray	1,164 (27.9)	313 (22.5)	804 (26.8)	275 (22.7)	894 (28.3)	373 (23.3)
MRI	1,403 (33.6)	482 (34.6)	1,012 (33.8)	456 (37.6)	1,120 (35.4)	545 (34.0)
Other§	310 (7.4)	98 (7.0)	240 (8.0)	91 (7.5)	246 (7.8)	111 (6.9)

Imaging findings*						
Disc degeneration	1,893 (45.4)	681 (48.9)	1,332 (44.5)	641 (52.8)	1,319 (41.7)	826 (51.5)
Facet joint degeneration	445 (10.7)	158 (11.3)	314 (10.5)	146 (12.0)	309 (9.8)	205 (12.8)
Scoliosis	238 (5.7)	69 (4.9)	153 (5.1)	62 (5.1)	167 (5.3)	93 (5.8)
Spondylolysis	23 (0.6)	9 (0.6)	13 (0.4)	7 (0.6)	17 (0.5)	12 (0.7)
Spondylolisthesis	162 (3.9)	60 (4.3)	109 (3.6)	57 (4.7)	106 (3.4)	75 (4.7)
Spinal stenosis	212 (5.1)	100 (7.2)	171 (5.7)	100 (8.2)	144 (4.6)	115 (7.2)
Disc protrusion	211 (5.1)	90 (6.5)	150 (5.0)	81 (6.7)	134 (4.2)	82 (5.1)
Disc herniation (extrusion)	1,105 (26.5)	452 (32.4)	842 (28.1)	413 (34.0)	857 (27.1)	503 (31.4)
Unspecific syndrome	573 (13.7)	200 (14.3)	402 (13.4)	184 (15.2)	341 (10.8)	227 (14.2)
Other findings	365 (8.7)	120 (8.6)	233 (7.8)	100 (8.2)	262 (8.3)	131 (8.2)
No findings	1,692 (40.5)	514 (36.9)	1242 (41.5)	426 (35.1)	1,390 (44.0)	558 (34.8)
Treatments						
Drugs*						
Analgesics	2,797 (67.0)	935 (67.1)	2,035 (67.9)	841 (69.3)	2,121 (67.1)	1,105 (68.9)
Nonsteroidal anti-inflammatory drugs	2,725 (65.3)	871 (62.5)	1,978 (66.0)	768 (63.3)	2,098 (66.4)	1,049 (65.4)
Steroids	391 (9.4)	123 (8.8)	315 (10.5)	123 (10.1)	319 (10.1)	145 (9.0)
Muscle relaxants	1,019 (24.4)	311 (22.3)	770 (25.7)	281 (23.2)	829 (26.2)	354 (22.1)
Opioids	132 (3.2)	66 (4.7)	96 (3.2)	54 (4.5)	99 (3.1)	86 (5.4)
Other	986 (23.6)	365 (26.2)	675 (22.5)	334 (27.5)	744 (23.5)	446 (27.8)
Nonpharmacological treatments*						
Physical therapy/ rehabilitation	580 (13.9)	201 (14.4)	454 (15.2)	172 (14.2)	441 (14.0)	226 (14.1)

MRI, magnetic resonance imaging; NDI, Neck Disability Index; NRT, neuroreflexotherapy; RP, severity of referred pain (in the 6,649 patients who had it); SP, severity of spinal pain; VAS, visual analog scale (range from better to worse; 0–10).

* Frequency (%).

† Median (percentile 25; percentile 75).

‡ Type of pain: “Radicular pain caused by disc protrusion/extrusion or spinal stenosis” if (a) severity of referred pain \geq spinal pain, (b) corresponding imaging finding on MRI, (c) distribution of pain consistent with the nerve root compressed by the corresponding imaging finding. “Nonspecific pain” if one or more of these criteria were not met.

§ Other diagnostic procedures: electromyography, computed tomography scan, and other.

|| Other imaging findings: annular tear, loss of cervical lordosis, loss of thoracic cifosis, loss of lumbar lordosis, horizontalization of the sacrum, lumbarization of S1, sacralization of L5.

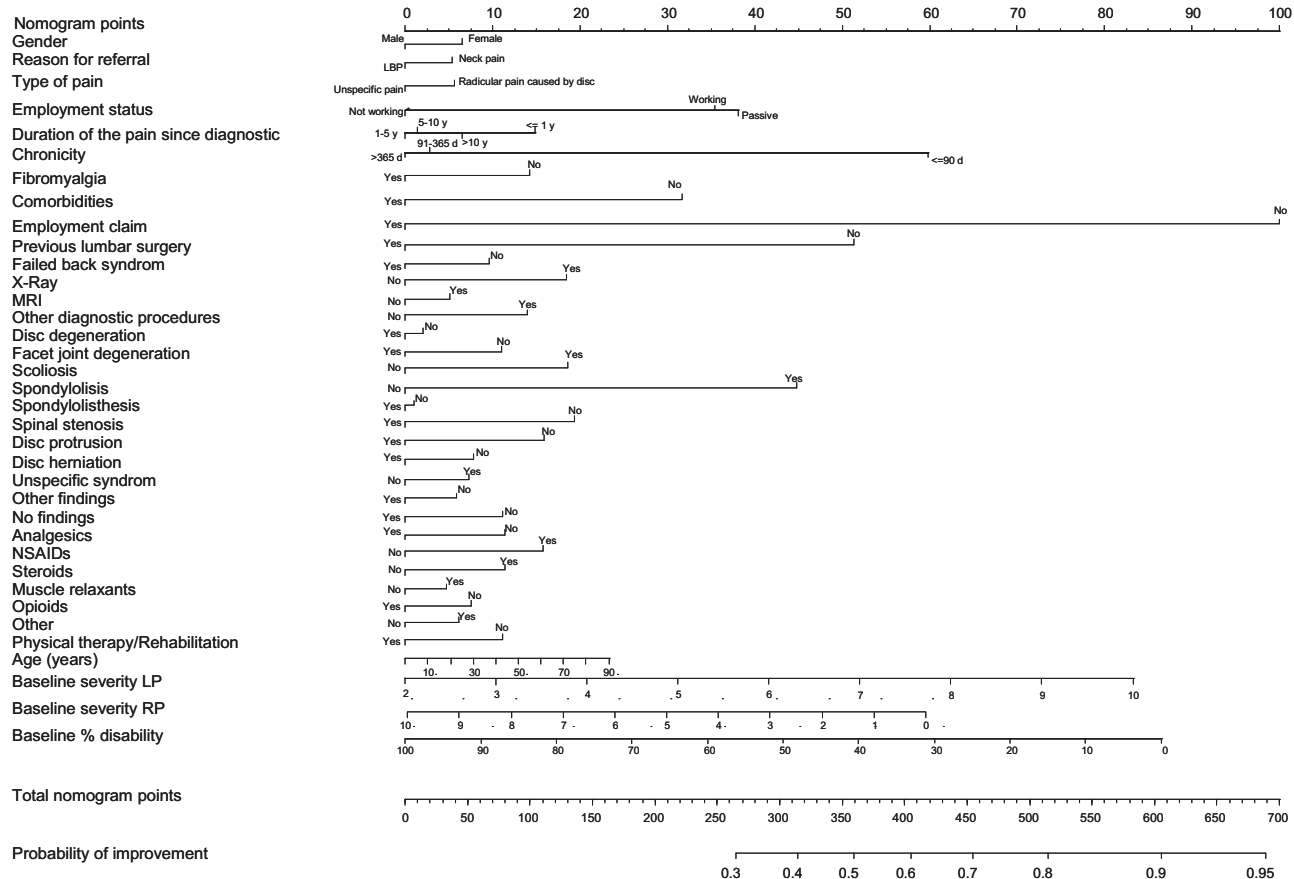


Fig. 2. Nomogram for improvement of spinal pain.

As a result, the “complete cases” analysis included 5,568 subjects for spinal pain, 4,209 for referred pain, and 4,764 for disability (Fig. 1). At discharge, a clinically relevant improvement was experienced by 4,174 (75%) patients for spinal pain, 2,996 patients (71.1%) for referred pain, and 3,160 (66.3%) for disability (Table 2).

Among the 2,588 patients treated for neck pain, 1,397 (54%) underwent NRT after the Spanish version of the NDI had become available. The volume of missing data ranged from 4 (0.04%) for age, to 1,096 (12.5%) for employment status, and 2,762 (31.5%) of the 8,778 patients showed one or more missing data.

Once multiple imputation was performed, the analysis included 7,976 patients for spinal pain, 5,974 for referred pain, and 5,640 for disability. The same variables were included in the multiple imputation analysis and in the regression models, except for the variables, which were eliminated from the models because a prevalence of <1% (being involved in litigation, having a difference in leg length and using codeine).

The tables showing the coefficients of the final models for spinal pain, referred pain and disability, for both the complete cases and the imputation analyses, are available online. Results from both analyses identified the same variables as those which have the greatest predictive value. Their coefficients were also very similar, although limits of the

95% confidence interval were generally narrower in the imputation analysis and coefficients for variables with smaller predictive value showed larger variations across analyses.

Figures 2–4 show the nomograms corresponding to the models on improvement of spinal pain, referred pain and disability, respectively. The variables contributing the most to predicting a clinically relevant improvement in all outcomes (spinal pain, referred pain and disability) are: pain being subacute (as opposed to chronic or highly chronic), not having previously undergone spine surgery, not receiving financial assistance for neck pain or LBP (vs. having a “passive” employment status), and having undergone X-ray. In addition, a lower baseline severity of spinal pain is associated with a worse prognosis for improvement of spinal pain and a better one for referred pain; a lower baseline severity of referred pain is associated with a worse prognosis for referred pain and a better one for spinal pain and disability, and a lower degree of disability is associated with a worse prognosis for disability and a better one for spinal pain and referred pain. Not having comorbidities is associated with a better prognosis for spinal pain and referred pain, but not for disability.

Additional variables contributing significantly to predicting a clinically relevant improvement in spinal pain are “not being involved in an employment claim” and “using NSAIDs”; for improvement in referred pain, “not

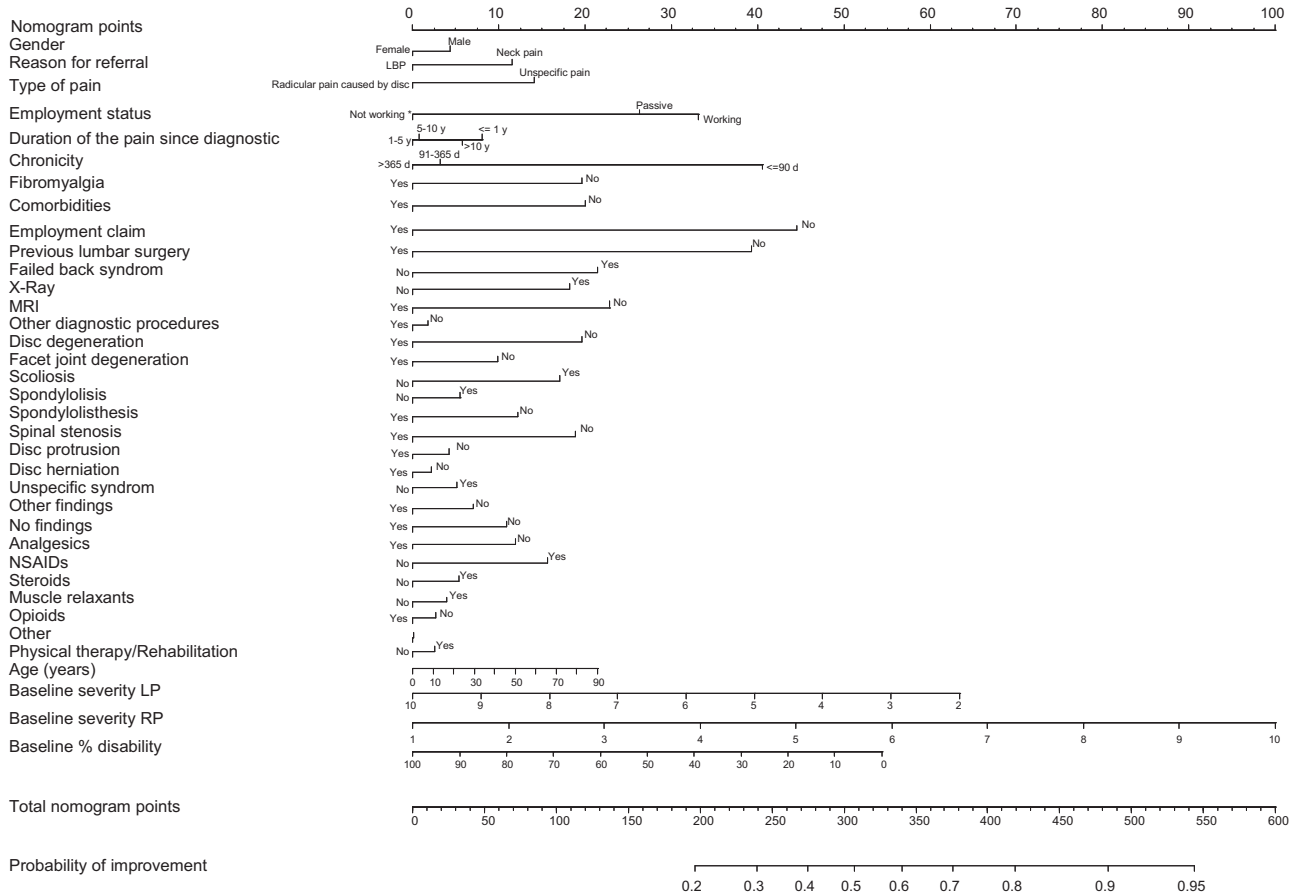


Fig. 3. Nomogram for improvement of referred pain.

showing disc degeneration” and “using NSAIDs”; and for improvement in disability, “having been treated for LBP (vs. for neck pain)” and “using muscle relaxants.”

Discrimination of the three models was slight to moderate [48], with a corrected D_{xy} index of 0.241 for spinal pain, 0.278 for referred pain, and 0.215 for disability. Calibration of the models was good, as shown in the three calibration plots available online. These plots show that the models accurately predict the actual likelihood of improvement in spinal pain, referred pain and disability, especially in the range of probabilities predicted by a greater number of actual observations.

Discussion

These results show that it is feasible to use registries and postimplementation surveillance mechanisms in routine practice to identify factors predicting the clinical response to treatments. The models developed in this study are valid for predicting a clinically relevant improvement in spinal pain, referred pain, and disability after undergoing a minimally invasive, nonpharmacological treatment. The nomograms based on these models make it possible to quantify the likelihood that a given patient will experience

a clinically relevant improvement in routine practice if undergoing this procedure (Figs. 2–4). For instance, according to these nomograms, the probability of experiencing a clinically relevant improvement in spinal pain after NRT is 87% for a 40-year old housewife who presents a lumbar disc herniation and no other findings on MRI, has been symptomatic for the first time in her life during the last 6 weeks, with scores of 8 VAS points for both spinal and referred pain, and 19 Roland-Morris points for disability, has no other comorbidities, has not undergone other diagnostic procedures or rehabilitation, and is using opioids, NSAIDs, and steroids. However, the probability would have been 72% if pain had lasted for more than 1 year and she had been receiving compensation benefits. The use of these methods to produce nomograms for all neck pain and LBP treatments would provide patients and clinicians with valuable tools in the context of informed shared decision making.

Registries compiled in routine practice, which include large numbers of variables and subjects, often present missing data [23,49]. Under the assumption of a missing at random pattern, performing multiple imputation analysis by chained equations is more efficient than restricting the analysis to complete cases [25,49,50]. Moreover, the former could address the potential risk of bias implied by the latter

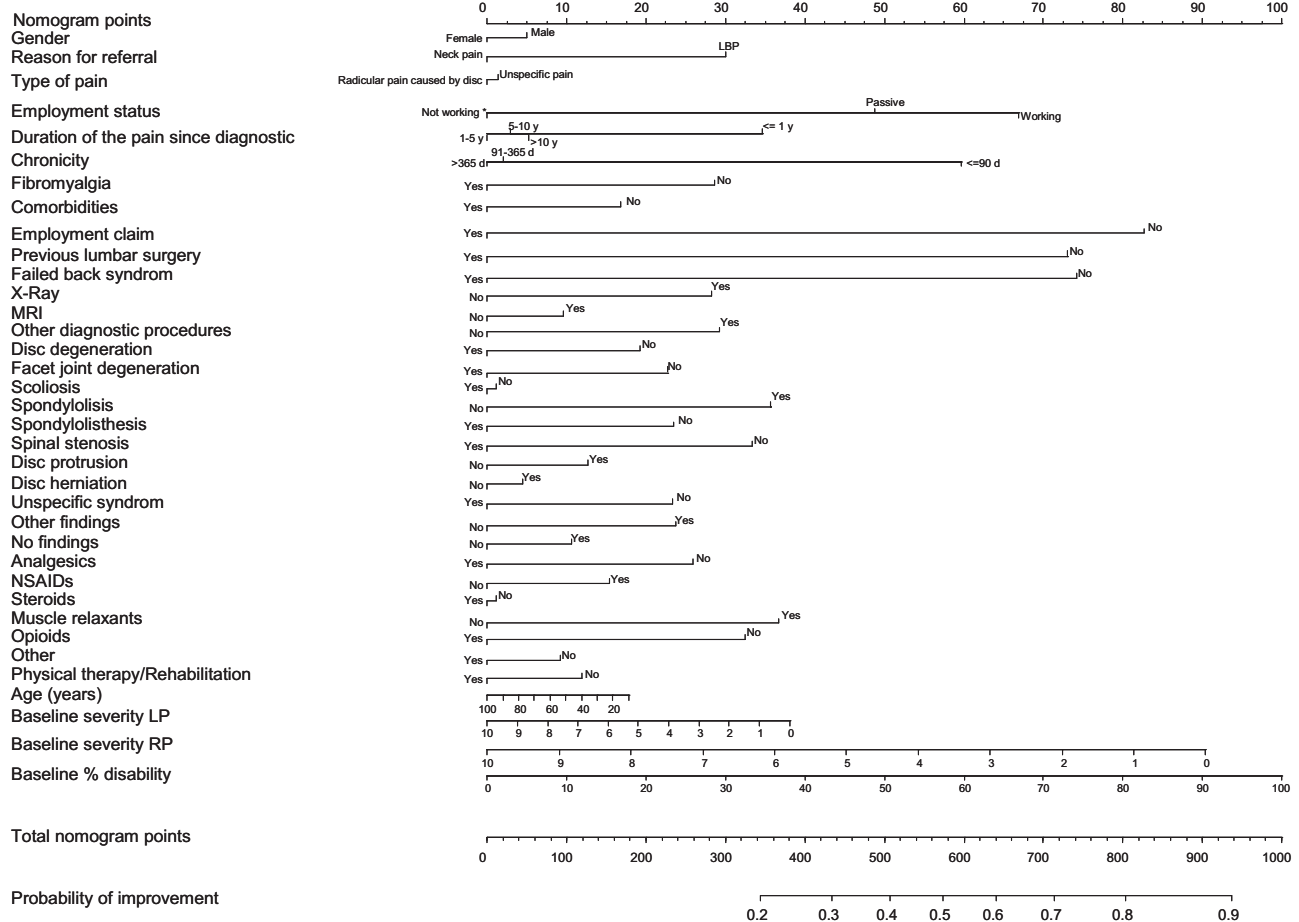


Fig. 4. Nomogram for improvement of disability.

[26,30]. The current study had approximately 30% of missing data, and multiple imputation led to a rise of between 20% and 40% in the number of cases analyzed in each model. This increased statistical power and made it possible to identify variables with a relevant prognostic value. Nevertheless, the coefficients of the variables with the greatest predictive value were very similar in both analyses, which supports their validity.

This registry included all SNHS patients who underwent a specific form of conservative treatment during a period of over 7 years across five different geographic regions. It gathered data on clinically relevant variables through previously validated methods [17,19,20,32,34,40]. The calibration and discrimination of the three predictive models were assessed using complex statistical and computational techniques, which provide bias-corrected indexes (resampling by bootstrap) [29,47], and take into account the multiple datasets derived from multiple imputation. These models had only slight-to-moderate discrimination capacity but showed good calibration, so the nomograms based on these models are applicable in routine practice.

However, a registry analysis is an observational study, and “association” does not necessarily imply “causation.” Therefore, results showing the association between a given

variable and a better or worse prognosis should be interpreted cautiously, taking clinical plausibility into account. For instance, prescribing X-rays has shown to increase patients’ satisfaction, but to be clinically superfluous in patients with neck pain or LBP who do not show “red flags” for potential systemic diseases [10,14,51]. Therefore, it is unlikely that undergoing X-rays “improves” patients’ prognosis; it is more likely that, when patients insist on having imaging, clinicians prescribe X-rays (as opposed to other more complex imaging procedures, such as magnetic resonance) to those in whom there is no reason to suspect any serious condition (ie, those for whom they expect a better prognosis). Similarly, taking into account evidence on the effectiveness of muscle relaxants [9,14,52], it is likely the association between using these drugs and a slightly better prognosis for disability reflects that clinicians prescribe them to patients who they perceive are less severely impaired.

Nevertheless, most associations found in this study are plausible, consistent with the existing evidence, and make clinical sense; NSAIDs are effective for treating SP [9,14,52]; receiving financial benefits for neck pain or LBP is associated with delayed recovery [53–57]; and the prognosis of LBP and neck pain is worse for patients

who have been in pain for longer, which also implies that previous treatments have failed [1,2,22,53]. Moreover, a greater baseline value for a given variable (spinal pain, referred pain, or disability) leaves more room for its improvement [58], either because of treatment or regression to the mean, whereas recovery is more difficult among patients who are more severely affected (in terms of the other variables). It also makes clinical sense that the prognosis of pain is worse for patients involved in employment claims [53–57], and among those with comorbidities (Figs. 2–4), suggesting poorer general health, less physically active lifestyles and potential additional psychological distress.

This study has some weaknesses. Academic level has been reported to influence LBP prognosis [53,59,60], but data on education level could not be included in this study because they were available only for a minority of patients. However, from May 2009 onwards these data have been systematically registered and will be analyzed in future reports. This registry does not include any psychological variables, whereas pain is a sensory and emotional experience which is influenced by psychological factors [53,59–68]. However, the psychological variables influencing the prognosis of patients with LBP specifically in the Spanish cultural setting are yet to be identified; to date, only the influence of fear avoidance beliefs and catastrophizing have been assessed, and shown to be clinically irrelevant or null [40,44,69–72]. Once psychological variables with a prognostic influence in this setting are identified, they will be added to this registry. The definition of the degree of improvement that is clinically important may vary across procedures and settings. However, the definitions used in this study are those which have shown to be valid in the setting where the study took place [37,38], and are in line with those determined in other settings [39]. Moreover, most patients were severely impaired and few had baseline scores below the value established as an MCIC (Table 1). Therefore, it is unlikely that using other definitions would have a major impact on results from this study.

In conclusion, this study developed three models that predict the evolution of spinal pain, referred pain, and disability in subacute and chronic patients treated with NRT for neck pain and LBP. Good calibration of these models suggests that they are applicable in routine practice, and nomograms allow clinicians and patients to quantify the likelihood of improvement for a given patient. Generalizing this approach to all forms of neck pain and LBP treatment could be valuable in the context of informed shared decision-making.

Appendix

Supplementary material

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.spinee.2013.09.039>.

References

- [1] Waddell G. *The back pain revolution*. London, UK: Churchill-Livingstone, 2004.
- [2] Andersson G. The epidemiology of spinal disorders. In: Frymoyer JW, ed. *The adult spine: principles and practice*. New York, NY: Raven Press, Ltd, 1997:93–141.
- [3] Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 2010;24:769–81.
- [4] Ivanova JI, Birnbaum HG, Schiller M, et al. Real-world practice patterns, health-care utilization, and costs in patients with low back pain: the long road to guideline-concordant care. *Spine J* 2011;11:622–32.
- [5] Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA* 2008;299:656–64.
- [6] van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. *Pain* 1995;62:233–40.
- [7] Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* 2008;8:8–20.
- [8] Gore M, Sadosky A, Stacey BR, et al. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and healthcare costs in usual care settings. *Spine* 2012;37:E668–77.
- [9] Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15(2 Suppl):S192–300.
- [10] Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet* 2009;373:463–72.
- [11] Haldeman S, Carroll L, Cassidy JD. Findings from the bone and joint decade 2000 to 2010 task force on neck pain and its associated disorders. *J Occup Environ Med* 2010;52:424–7.
- [12] Hancock MJ, Maher CG, Latimer J, et al. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *Eur Spine J* 2007;16:1539–50.
- [13] Kleinstuck F, Dvorak J, Mannion AF. Are “structural abnormalities” on magnetic resonance imaging a contraindication to the successful conservative treatment of chronic nonspecific low back pain? *Spine* 2006;31:2250–7.
- [14] Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478–91.
- [15] Urrutia G, Burton K, Morral A, et al. Neuroreflexotherapy for non-specific low back pain: a systematic review. *Spine* 2011;30:E148–53.
- [16] Kovacs FM, Abaira V, Pozo F, et al. Local and remote sustained trigger point therapy for exacerbations of chronic low back pain. A randomized, double-blind, controlled, multicenter trial. *Spine* 1997;22:786–97.
- [17] Kovacs FM, Llobera J, Abaira V, et al. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice: a cluster randomized, controlled trial. *Spine* 2002;27:1149–59.
- [18] Kovacs FM, Abaira V, Lopez-Abente G, Pozo F. [Neuro-reflexotherapy intervention in the treatment of non specified low back pain: a randomized, controlled, double-blind clinical trial]. *Med Clin (Barc)* 1993;101:570–5.
- [19] Corcoll J, Orfila J, Tobajas P, Alegre L. Implementation of neuroreflexotherapy for subacute and chronic neck and back pain within the Spanish public health system: audit results after one year. *Health Policy* 2006;79:345–57.
- [20] Grupo PINS. [Implementation of neuroreflexotherapy in the treatment of back pain. Results of a pilot trial]. *Gac Sanit* 2004;18:275–86.
- [21] Pallicer A, Corcoll J, Orfila J. Implementation and post-marketing surveillance of neuroreflexotherapy in the routine practice of the

- Spanish Public Health System. IX International Forum on Low Back Pain Research, Palma de Mallorca, October 2007. Available at: <http://www.aemen.es/descargas/Implementationandpostmarketing-surveillanceofneuroreflexotherapy.pdf> Accessed September 8, 2008.
- [22] Kovacs F, Abaira V, Muriel A, et al. Prognostic factors for neuro-reflexotherapy in the treatment of subacute and chronic neck and back pain: a study of predictors of clinical outcome in routine practice of the Spanish National Health Service. *Spine* 2007;32:1621–8.
 - [23] White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30:377–99.
 - [24] Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20:40–9.
 - [25] van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.
 - [26] Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
 - [27] Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453–73.
 - [28] Mallett S, Royston P, Waters R, et al. Reporting performance of prognostic models in cancer: a review. *BMC Med* 2010;8:21.
 - [29] Harrell F. Regression modeling strategies with applications to linear models, logistics regression, and survival analysis. New York, NY: Springer, 2001.
 - [30] Vergouw D, Heymans MW, Peat GM, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Med Res Methodol* 2010;10:81.
 - [31] Instituto Nacional de Estadística. Cifras de población referidas al 01/01/2010; Instituto Nacional de Estadística. Available at: <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe260%2Fa2010%2F&file=pcaxis&N=&L=0>. Accessed November 5, 2013.
 - [32] Huskisson EC. Measurement of pain. *Lancet* 1974;2:1127–31.
 - [33] AEMEN-Asociación Española de Médicos Neuroreflexoterapeutas [Spanish association of neuro-reflexotherapy physicians]. Available at: <http://www.aemen.es/formacion.html>. Accessed November 5, 2013.
 - [34] Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998;23:2003–13.
 - [35] Kovacs FM, Llobera J, Gil Del Real MT, et al. Validation of the Spanish version of the Roland-Morris questionnaire. *Spine* 2002;27:538–42.
 - [36] Kovacs FM, Bago J, Royuela A, et al. Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. *BMC Musculoskelet Disord* 2008;9:42.
 - [37] Kovacs FM, Abaira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with non-specific low back pain. *Spine* 2007;32:2915–20.
 - [38] Kovacs FM, Abaira V, Royuela A, et al. Minimum detectable and minimal clinically important changes for pain in patients with non-specific neck pain. *BMC Musculoskelet Disord* 2008;9:43.
 - [39] Pool JJ, Ostelo RW, Hoving JL, et al. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. *Spine* 2007;32:3047–51.
 - [40] Kovacs F, Noguera J, Abaira V, et al. The influence of psychological factors on low back pain-related disability in community dwelling older persons. *Pain Med* 2008;9:871–80.
 - [41] Barrett TW, Martin AR, Storrow AB, et al. A clinical prediction model to estimate risk for 30-day adverse events in emergency department patients with symptomatic atrial fibrillation. *Ann Emerg Med* 2011;57:1–12.
 - [42] Heymans MW, van BS, Knol DL, et al. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. *Spine J* 2010;10:847–56.
 - [43] IASP. Classification of chronic pain: description of chronic pain syndromes and definitions of pain terms. Seattle, WA: IASP Press, 1994.
 - [44] Kovacs FM, Abaira V, Zamora J, Fernandez C. The transition from acute to subacute and chronic low back pain: a study based on determinants of quality of life and prediction of chronic disability. *Spine* 2005;30:1786–92.
 - [45] Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* 2007;8:206–13.
 - [46] Rubin D. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley & Sons, 1987.
 - [47] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
 - [48] Turner PL, Saager L, Dalton J, et al. A nomogram for predicting surgical complications in bariatric surgery patients. *Obes Surg* 2011;21:655–62.
 - [49] White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med* 2010;29:2920–31.
 - [50] Janssen KJ, Donders AR, Harrell FE Jr, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol* 2010;63:721–7.
 - [51] Deyo RA. Imaging idolatry: the uneasy intersection of patient satisfaction, quality of care, and overuse. *Arch Intern Med* 2009;169:921–3.
 - [52] Kuijpers T, van MM, Rubinstein SM, et al. A systematic review on the effectiveness of pharmacological interventions for chronic non-specific low-back pain. *Eur Spine J* 2011;20:40–50.
 - [53] Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA* 2010;303:1295–302.
 - [54] Rasmussen C, Leboeuf-Yde C, Hestbaek L, Manniche C. Poor outcome in patients with spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in the secondary care sector. *Scand J Rheumatol* 2008;37:462–8.
 - [55] Scuderi GJ, Sherman AL, Brusovanik GV, et al. Symptomatic cervical disc herniation following a motor vehicle collision: return to work comparative study of workers' compensation versus personal injury insurance status. *Spine J* 2005;5:639–44.
 - [56] Rainville J, Sobel JB, Hartigan C, Wright A. The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. *Spine* 1997;22:2016–24.
 - [57] Atlas SJ, Chang Y, Kammann E, et al. Long-term disability and return to work among patients who have a herniated lumbar disc: the effect of disability compensation. *J Bone Joint Surg Am* 2000;82:4–15.
 - [58] Helmhout PH, Staal JB, Heymans MW, et al. Prognostic factors for perceived recovery or functional improvement in non-specific low back pain: secondary analyses of three randomized clinical trials. *Eur Spine J* 2010;19:650–9.
 - [59] Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *J Clin Epidemiol* 2009;62:781–96.
 - [60] Lakke SE, Soer R, Takken T, Reneman MF. Risk and prognostic factors for non-specific musculoskeletal pain: a synthesis of evidence from systematic reviews classified into ICF dimensions. *Pain* 2009;147:153–64.
 - [61] Carragee EJ, Alamin TF, Miller JL, Carragee JM. Discographic, MRI and psychosocial determinants of low back pain disability and remission: a prospective study in subjects with benign persistent back pain. *Spine J* 2005;5:24–35.

- [62] Grotle M, Foster NE, Dunn KM, Croft P. Are prognostic indicators for poor outcome different for acute and chronic low back pain consultants in primary care? *Pain* 2010;151:790–7.
- [63] Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 2002;27:E109–20.
- [64] Smeets RJ, Vlaeyen JW, Kester AD, Knottnerus JA. Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *J Pain* 2006;7:261–71.
- [65] Smeets RJ, Maher CG, Nicholas MK, et al. Do psychological characteristics predict response to exercise and advice for subacute low back pain? *Arthritis Rheum* 2009;61:1202–9.
- [66] Spinhoven P, Ter KM, Koe-Snijders AM, et al. Catastrophizing and internal pain control as mediators of outcome in the multidisciplinary treatment of chronic low back pain. *Eur J Pain* 2004;8: 211–9.
- [67] Truchon M, Cote D. Predictive validity of the Chronic Pain Coping Inventory in subacute low back pain. *Pain* 2005;116:205–12.
- [68] Wand BM, McAuley JH, Marston L, De Souza LH. Predicting outcome in acute low back pain using different models of patient profiling. *Spine* 2009;34:1970–5.
- [69] Albaladejo C, Kovacs FM, Royuela A, et al. The efficacy of a short education program and a short physiotherapy program for treating low back pain in primary care: a cluster randomized trial. *Spine* 2010;35:483–96.
- [70] Kovacs F, Abaira V, Cano A, et al. Fear avoidance beliefs do not influence disability and quality of life in Spanish elderly subjects with low back pain. *Spine* 2007;32:2133–8.
- [71] Kovacs F, Abaira V, Santos S, et al. A comparison of two short education programs for improving low back pain-related disability in the elderly: a cluster randomized controlled trial. *Spine* 2007;32: 1053–9.
- [72] Kovacs FM, Seco J, Royuela A, et al. The prognostic value of catastrophizing for predicting the clinical evolution of low back pain patients: a study in routine clinical practice within the Spanish National Health Service. *Spine J* 2012;22:545–55.

Supplemental Table 1. Predictive model for the improvement of SP.

Variables	Complete cases analysis (n=5568)		Multiple imputation analysis (n= 7976)	
	Coefficients	p	Coefficients	p
Gender (female)	0.13 (-0.01; 0.27)	0.065	0.06 (-0.06; 0.18)	0.340
Reason for referral to NRT (LBP)	0.07 (-0.12; 0.26)	0.478	-0.05 (-0.19; 0.09)	0.490
Type of pain (Radicular pain caused by disc)	0.12 (-0.18; 0.42)	0.440	0.05 (-0.20; 0.31)	0.698
Employment status				
Passive	Ref. cat.		Ref. cat.	
Receiving financial assistance for NP or LBP	-0.30 (-0.53; -0.06)	0.012	-0.34 (-0.56; -0.12)	0.003
Working	0.01 (-0.16; 0.17)	0.953	-0.02 (-0.17; 0.13)	0.760
Duration of the pain since diagnostic categorized				
≤ 1 year	Ref. cat.		Ref. cat.	
1-5 years	-0.14 (-0.34; 0.07)	0.187	-0.13 (-0.30; 0.04)	0.127
5-10 years	-0.07 (-0.29; 0.14)	0.491	-0.12 (-0.30; 0.06)	0.182
>10 years	-0.03 (-0.25; 0.18)	0.768	-0.07 (-0.26; 0.11)	0.420
Duration of the pain episode (days) categorized				
Subacute (≤ 90 days)	Ref. cat.		Ref. cat.	
Chronic (91-365 days)	-0.50 (-0.66; -0.33)	0.000	-0.51 (-0.65; -0.37)	0.000
Highly Chronic (>365 days)	-0.58 (-0.76; -0.39)	0.000	-0.53 (-0.69; -0.38)	0.000
Concomitant diagnosis of fibromyalgia	-0.07 (-0.42; 0.27)	0.676	-0.13 (-0.41; 0.16)	0.379
Other comorbidities	-0.25 (-0.43; -0.08)	0.004	-0.28 (-0.43; -0.14)	0.000
Employment claim	-0.93 (-1.60; -0.26)	0.006	-0.89 (-1.44; -0.35)	0.001
Previous lumbar surgery	-0.54 (-0.76; -0.31)	0.000	-0.46 (-0.65; -0.26)	0.000
Failed back syndrome (yes)	-0.01 (-0.56; 0.54)	0.972	-0.09 (-0.58; 0.41)	0.732
X-Ray performed	0.20 (0.04; 0.35)	0.013	0.16 (0.03; 0.30)	0.014
MRI performed	0.14 (-0.01; 0.29)	0.074	0.05 (-0.08; 0.17)	0.483
Other diagnostic procedures performed	0.22 (-0.03; 0.47)	0.086	0.12 (-0.09; 0.34)	0.261
Imaging findings				
Disc degeneration	-0.03 (-0.24; 0.18)	0.781	-0.02 (-0.19; 0.16)	0.838
Facet joint degeneration	-0.03 (-0.24; 0.19)	0.809	-0.10 (-0.28; 0.08)	0.286
Scoliosis	0.08 (-0.22; 0.37)	0.611	0.17 (-0.08; 0.41)	0.180
Spondylolisthesis	0.22 (-0.28; 0.72)	0.394	0.40 (-0.06; 0.86)	0.087
Spondylolisthesis	-0.03 (-0.36; 0.30)	0.859	-0.01 (-0.28; 0.26)	0.945
Spinal stenosis	-0.15 (-0.42; 0.12)	0.282	-0.17 (-0.40; 0.05)	0.134
Disc protrusion	-0.03 (-0.31; 0.26)	0.854	-0.14 (-0.35; 0.06)	0.170
Disc herniation (extrusion)	-0.17 (-0.35; 0.01)	0.071	-0.07 (-0.22; 0.08)	0.364
Unspecific syndrome	0.02 (-0.18; 0.23)	0.834	0.06 (-0.10; 0.23)	0.435
Other findings	-0.07 (-0.31; 0.16)	0.546	-0.05 (-0.25; 0.15)	0.606
No findings	-0.22 (-0.48; 0.04)	0.095	-0.10 (-0.32; 0.12)	0.372
Analgesics	-0.15 (-0.31; 0.02)	0.075	-0.10 (-0.24; 0.04)	0.156
NSAIDs	0.19 (0.03; 0.36)	0.021	0.14 (0.002; 0.279)	0.047
Steroids	0.05 (-0.19; 0.29)	0.666	0.10 (-0.11; 0.31)	0.348
Muscle relaxants	0.11 (-0.06; 0.28)	0.194	0.04 (-0.10; 0.19)	0.571
Opioids	-0.18 (-0.51; 0.14)	0.275	-0.07 (-0.33; 0.20)	0.621
Other	0.07 (-0.10; 0.23)	0.427	0.05 (-0.08; 0.19)	0.445
Physical therapy/Rehabilitation	-0.11 (-0.30; 0.09)	0.283	-0.10 (-0.26; 0.07)	0.239
Age (years)	0.003 (-0.002; 0.010)	0.229	0.002 (-0.003; 0.007)	0.350
Baseline severity of SP	0.11 (0.07; 0.15)	0.000	0.09 (0.06; 0.13)	0.000
Baseline severity of RP	-0.07 (-0.09; -0.04)	0.000	-0.05 (-0.07; -0.03)	0.000
Baseline percentage of disability	-0.007 (-0.011; -0.005)	0.000	-0.01 (-0.011; -0.004)	0.000
Constant term	1.42 (0.93; 1.90)	0.000	1.72 (1.30; 2.13)	0.000

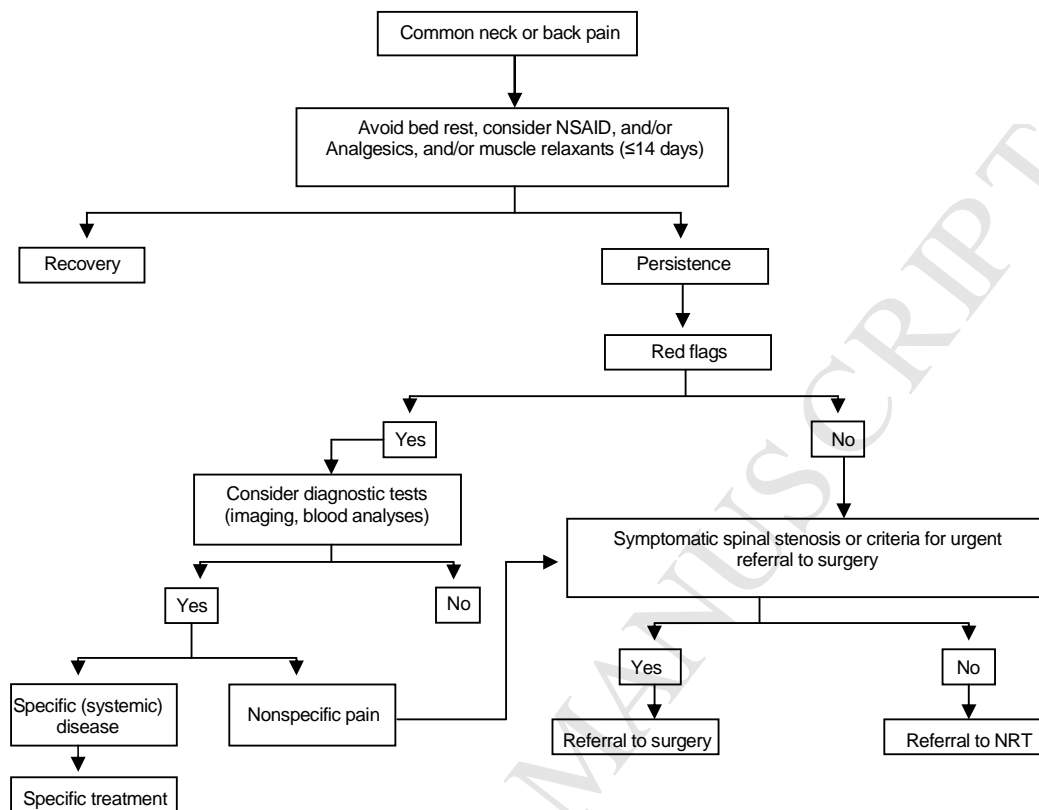
Supplemental Table 2. Predictive model for improvement of RP

Variables	Complete cases analysis (n=4209)		Multiple imputation analysis (n=5974)	
	Coefficients	p	Coefficients	p
Gender (female)	-0.08 (-0.23; 0.08)	0.318	-0.06 (-0.20; 0.08)	0.423
Reason for referral to NRT (LBP)	0.01 (-0.20; 0.22)	0.934	-0.14 (-0.30; 0.02)	0.077
Type of pain (Radicular pain caused by disc)	-0.16 (-0.48; 0.16)	0.315	-0.18 (-0.47; 0.10)	0.209
Employment status				
Passive	Ref. cat.			
Receiving financial assistance for NP or LBP	-0.28 (-0.54; -0.03)	0.028	-0.33 (-0.59; -0.08)	0.012
Working	0.08 (-0.10; 0.26)	0.381	0.08 (-0.09; 0.24)	0.362
Duration of the pain since diagnostic categorized				
≤ 1 year	Ref. cat.			
1-5 years	-0.07 (-0.30; 0.15)	0.531	-0.10 (-0.30; 0.10)	0.330
5-10 years	0.03 (-0.20; 0.27)	0.794	-0.09 (-0.31; 0.13)	0.407
>10 years	0.08 (-0.16; 0.32)	0.516	-0.03 (-0.25; 0.19)	0.804
Duration of the pain episode (days) categorized				
Subacute (≤ 90 days)	Ref. cat.			
Chronic (91-365 days)	-0.53 (-0.72; -0.35)	0.000	-0.47 (-0.63; -0.31)	0.000
Highly Chronic (>365 days)	-0.62 (-0.83; -0.42)	0.000	-0.51 (-0.69; -0.32)	0.000
Concomitant diagnosis of fibromyalgia	-0.11 (-0.47; 0.25)	0.539	-0.25 (-0.57; 0.07)	0.120
Other comorbidities	-0.18 (-0.37; 0.01)	0.058	-0.25 (-0.43; -0.07)	0.007
Employment claim	-0.39 (-1.13; 0.34)	0.293	-0.57 (-1.16; 0.02)	0.056
Previous lumbar surgery	-0.60 (-0.84; -0.36)	0.000	-0.49 (-0.72; -0.27)	0.000
Failed back syndrome (yes)	0.43 (-0.21; 1.06)	0.183	0.27 (-0.30; 0.84)	0.361
X-Ray performed	0.17 (-0.01; 0.34)	0.059	0.23 (0.01; 0.46)	0.043
MRI performed	-0.10 (-0.27; 0.06)	0.225	-0.29 (-0.57; -0.01)	0.041
Other diagnostic procedures performed	0.17 (-0.09; 0.44)	0.205	-0.02 (-0.35; 0.31)	0.893
Imaging findings				
Disc degeneration	-0.34 (-0.57; -0.10)	0.005	-0.25 (-0.46; -0.04)	0.019
Facet joint degeneration	-0.04 (-0.27; 0.20)	0.766	-0.12 (-0.34; 0.09)	0.259
Scoliosis	0.04 (-0.29; 0.36)	0.816	0.21 (-0.07; 0.49)	0.135
Spondylolysis	0.10 (-0.44; 0.63)	0.724	0.06 (-0.43; 0.56)	0.802
Spondylolisthesis	-0.11 (-0.47; 0.24)	0.527	-0.15 (-0.45; 0.15)	0.317
Spinal stenosis	-0.18 (-0.46; 0.11)	0.221	-0.24 (-0.48; 0.01)	0.056
Disc protrusion	0.05 (-0.27; 0.36)	0.765	-0.05 (-0.30; 0.19)	0.658
Disc herniation (extrusion)	-0.07 (-0.28; 0.13)	0.468	-0.03 (-0.20; 0.14)	0.735
Unspecific syndrome	-0.05 (-0.27; 0.18)	0.688	0.06 (-0.13; 0.25)	0.535
Other findings	-0.03 (-0.30; 0.24)	0.809	-0.09 (-0.32; 0.14)	0.437
No findings	-0.35 (-0.64; -0.05)	0.022	-0.14 (-0.42; 0.15)	0.348
Analgesics	-0.23 (-0.42; -0.04)	0.015	-0.15 (-0.32; 0.02)	0.084
NSAIDs	0.25 (0.06; 0.43)	0.009	0.19 (0.03; 0.36)	0.024
Steroids	-0.06 (-0.31; 0.20)	0.664	0.07 (-0.16; 0.30)	0.554
Muscle relaxants	0.13 (-0.06; 0.31)	0.185	0.05 (-0.12; 0.22)	0.563
Opioids	-0.02 (-0.39; 0.34)	0.899	-0.02 (-0.30; 0.25)	0.864
Other	-0.03 (-0.20; 0.15)	0.770	-0.00 (-0.16; 0.16)	0.998
Physical therapy/Rehabilitation	0.09 (-0.12; 0.30)	0.411	0.03 (-0.15; 0.22)	0.724
Age (years)	0.004 (-0.002; 0.010)	0.206	0.003 (-0.004; 0.009)	0.391
Baseline severity of SP	-0.07 (-0.12; -0.02)	0.004	-0.10 (-0.14; -0.06)	0.000
Baseline severity of RP	0.11 (0.07; 0.16)	0.000	0.14 (0.10; 0.18)	0.000
Baseline percentage of disability	-0.007 (-0.011; -0.004)	0.000	-0.007(-0.010; -0.004)	0.000
Constant term	1.77 (1.22; 2.33)	0.000	2.09 (1.47; 2.71)	0.000

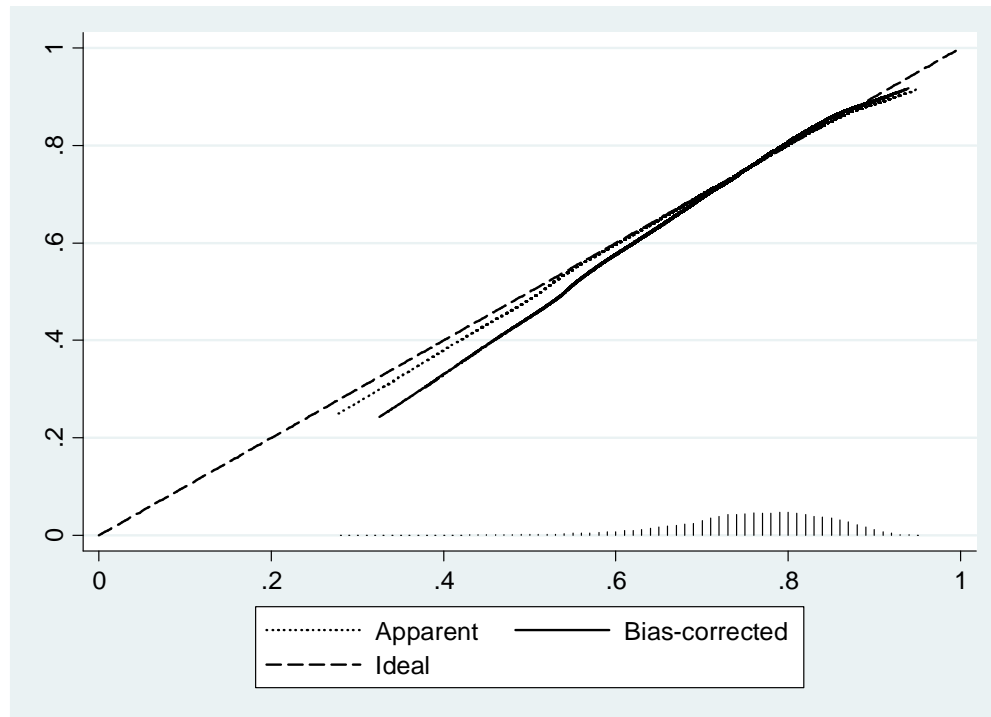
Supplemental Table 3. Predictive model for improvement of disability

Variables	Complete cases analysis (n=4764)		Multiple imputation analysis (n=5640)	
	Coefficients	p	Coefficients	p
Gender (female)	-0.02 (-0.15; 0.12)	0.811	-0.03 (-0.16; 0.10)	0.650
Reason for referral to NRT (LBP)	0.23 (0.02; 0.43)	0.028	0.18 (0.01; 0.36)	0.045
Type of pain (Radicular pain caused by disc)	-0.09 (-0.39; 0.22)	0.577	-0.01 (-0.28; 0.26)	0.954
Employment status				
Passive	Ref. cat.		Ref. cat.	
Receiving financial assistance for NP or LBP	-0.26 (-0.50; -0.03)	0.027	-0.29 (-0.52; -0.07)	0.012
Working	0.11 (-0.05; 0.28)	0.169	0.11 (-0.05; 0.27)	0.179
Duration of the pain since diagnostic categorized				
≤ 1 year	Ref. cat.		Ref. cat.	
1-5 years	-0.13 (-0.33; 0.07)	0.215	-0.19 (-0.38; -0.01)	0.045
5-10 years	-0.12 (-0.33; 0.09)	0.270	-0.21 (-0.40; -0.02)	0.031
>10 years	-0.18 (-0.39; 0.03)	0.096	-0.18 (-0.37; 0.02)	0.073
Duration of the pain episode (days) categorized				
Subacute (≤ 90 days)	Ref. cat.		Ref. cat.	
Chronic (91-365 days)	-0.38 (-0.54; -0.23)	0.000	-0.35 (-0.49; -0.20)	0.000
Highly Chronic (>365 days)	-0.40 (-0.58; -0.23)	0.000	-0.36 (-0.52; -0.20)	0.000
Concomitant diagnosis of fibromyalgia	-0.19 (-0.53; 0.16)	0.289	-0.17 (-0.49; 0.15)	0.288
Other comorbidities	-0.10 (-0.27; 0.069)	0.246	-0.10 (-0.25; 0.05)	0.188
Employment claim	-0.39 (-1.09; 0.31)	0.275	-0.50 (-1.09; 0.09)	0.097
Previous lumbar surgery	-0.50 (-0.72; -0.27)	0.000	-0.44 (-0.65; -0.23)	0.000
Failed back syndrome (yes)	-0.44 (-1.00; 0.12)	0.125	-0.45 (-0.96; 0.06)	0.087
X-Ray performed	0.20 (0.06; 0.35)	0.009	0.17 (0.03; 0.31)	0.017
MRI performed	0.09 (-0.06; 0.24)	0.231	0.06 (-0.08; 0.20)	0.411
Other diagnostic procedures performed	0.21 (-0.04; 0.45)	0.099	0.18 (-0.05; 0.40)	0.131
Imaging findings				
Disc degeneration	-0.19 (-0.40; 0.02)	0.074	-0.12 (-0.30; 0.07)	0.215
Facet joint degeneration	-0.14 (-0.35; 0.07)	0.186	-0.14 (-0.32; 0.05)	0.151
Scoliosis	0.01 (-0.27; 0.29)	0.964	-0.01 (-0.25; 0.24)	0.955
Spondylolisthesis	0.15 (-0.31; 0.62)	0.514	0.22 (-0.20; 0.63)	0.312
Spondylolisthesis	-0.20 (-0.52; 0.12)	0.226	-0.14 (-0.42; 0.14)	0.327
Spinal stenosis	-0.15 (-0.43; 0.13)	0.285	-0.20 (-0.44; 0.04)	0.106
Disc protrusion	0.15 (-0.16; 0.46)	0.342	0.08 (-0.17; 0.32)	0.544
Disc herniation (extrusion)	0.01 (-0.17; 0.19)	0.937	0.03 (-0.13; 0.19)	0.741
Unspecific syndrome	-0.20 (-0.41; 0.01)	0.064	-0.14 (-0.32; 0.04)	0.119
Other findings	0.09 (-0.15; 0.33)	0.448	0.14 (-0.07; 0.36)	0.188
No findings	-0.01 (-0.27; 0.24)	0.915	0.06 (-0.17; 0.30)	0.588
Analgesics	-0.14 (-0.30; 0.02)	0.092	-0.16 (-0.31; -0.01)	0.042
NSAIDs	0.06 (-0.10; 0.22)	0.457	0.09 (-0.06; 0.24)	0.220
Steroids	-0.01 (-0.25; 0.22)	0.927	-0.01 (-0.23; 0.21)	0.951
Muscle relaxants	0.25 (0.08; 0.41)	0.003	0.22 (0.07; 0.37)	0.004
Opioids	-0.30 (-0.62; 0.01)	0.059	-0.20 (-0.47; 0.08)	0.156
Other	-0.03 (-0.19; 0.13)	0.753	-0.06 (-0.20; 0.09)	0.447
Physical therapy/Rehabilitation	-0.08 (-0.28; 0.11)	0.389	-0.07 (-0.25; 0.11)	0.425
Age (years)	0.001 (-0.005; 0.007)	0.711	-0.001 (-0.01; 0.00)	0.654
Baseline severity of SP	-0.02 (-0.06; 0.02)	0.380	-0.02 (-0.06; 0.01)	0.201
Baseline severity of RP	-0.06 (-0.08; -0.04)	0.000	-0.05 (-0.07; -0.03)	0.000
Baseline percentage of disability	0.006 (0.003; 0.009)	0.000	0.006 (0.003; 0.009)	0.000
Constant term	1.07 (0.58; 1.56)	0.000	1.23 (0.77; 1.69)	0.000

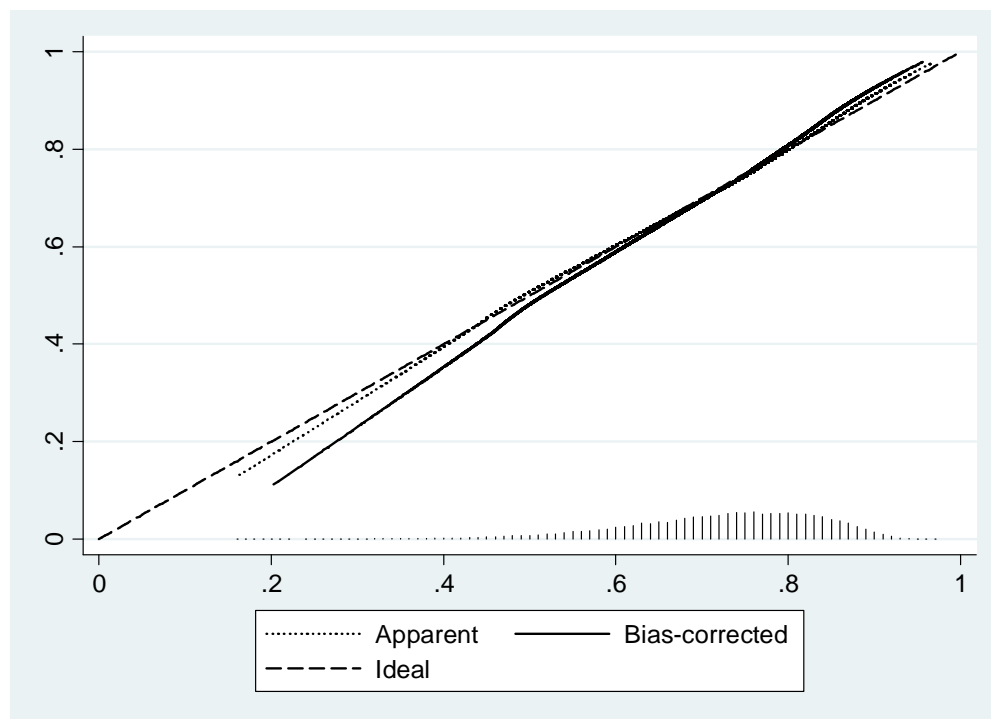
Supplemental figure 1. Referral protocol to NRT, within the Spanish National Health Service.



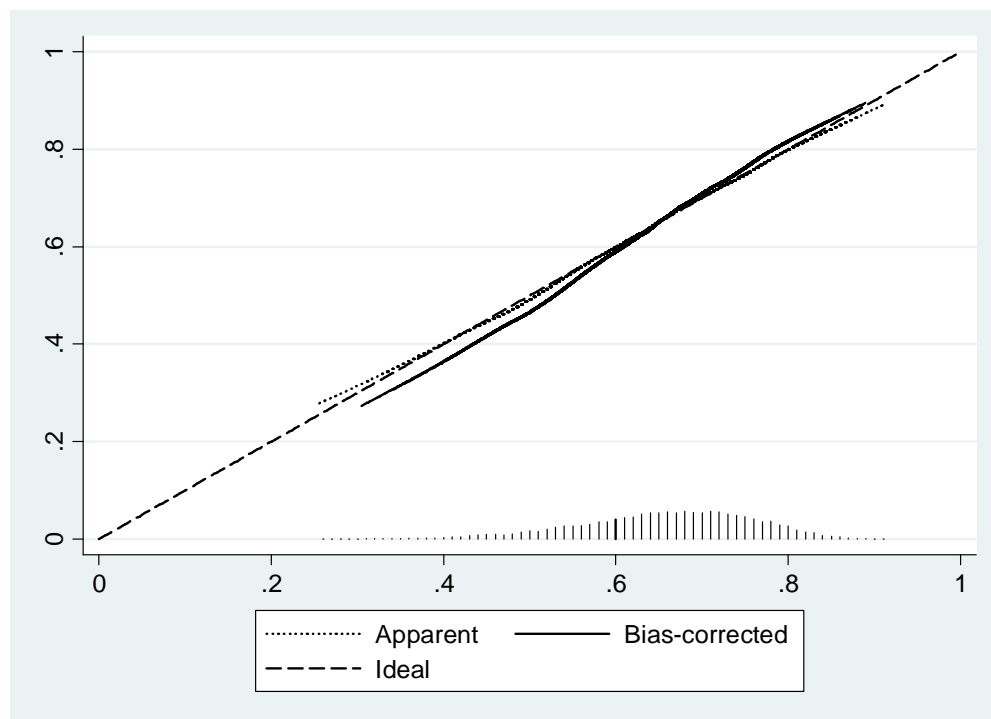
Supplemental Figure 2. Bootstrap calibration curve for the SP logistic model (500 bootstrap replications). The rug plot at the bottom of the graph indicates the distribution of predicted probabilities of spinal pain improvement.



Supplemental Figure 3. Bootstrap calibration curve for the RP logistic model (500 bootstrap replications). The rug plot at the bottom of the graph indicates the distribution of predicted probabilities of referred pain improvement.



Supplemental Figure 4. Bootstrap calibration curve for the disability logistic model (500 bootstrap replications). The rug plot at the bottom of the graph indicates the distribution of predicted probabilities of disability improvement.



DISCUSIÓN

Esta Tesis Doctoral muestra una aproximación a la labor de facilitar la aplicación de los resultados de la investigación científica para fundamentar la práctica clínica, empleando métodos estadísticos avanzados con el fin de conocer de forma válida las variables asociadas al pronóstico del dolor lumbar y cervical en dos cohortes de pacientes.

Con el fin de rellenar el hueco existente entre la investigación clínica y la aplicación de los resultados a la práctica diaria, se han desarrollado dos herramientas web que hacen accesibles los resultados de esta investigación. Además, se está trabajando en el desarrollo de una aplicación para teléfonos móviles inteligentes.

Los estudios pronósticos pueden centrarse sobre una cohorte de pacientes para estudiar el pronóstico basal sin haber recibido tratamientos modificadores, o también examinar factores predictores del pronóstico en una serie de pacientes que hayan recibido un tratamiento determinado (3). En esta Tesis, se han abordado ambos enfoques. En el primer artículo, el objetivo era obtener factores pronósticos asociados a pacientes con lumbalgia inespecífica, dadas sus características basales, tratamientos recibidos y pruebas diagnósticas prescritas antes o durante la visita basal. En el segundo, la entrada al estudio se determinó por la intervención NRT para todos los participantes y el objetivo era conocer su pronóstico a los 3 meses de dicha intervención.

Moons et al.,(3) recomiendan que los factores predictores analizados en un modelo pronóstico deberían estar claramente definidos, estandarizados y ser reproducibles, para así garantizar la generalización y aplicabilidad de los resultados del estudio a la práctica clínica. Apunta con especial interés hacia los factores predictores que requieren una interpretación subjetiva, como por ejemplo resultados de pruebas de imagen, ya que hay riesgo de acabar evaluando la capacidad predictiva del observador más que los propios predictores. Además, aconseja que los factores predictores debieran ser medidos con métodos aplicables o potencialmente aplicables a la práctica médica rutinaria, ya que técnicas de medición muy especializadas, pueden desencadenar predicciones sobre-optimistas. En esta Tesis, se han utilizado instrumentos validados y adaptados a la cultura española, encontrándose su uso

ampliamente extendido en la práctica médica rutinaria. Las variables resultado, como la escala visual analógica en la intensidad del dolor, el cuestionario Roland Morris en la discapacidad asociada a dolor lumbar y el Índice de Discapacidad Cervical en la discapacidad asociada a dolor cervical, son instrumentos que se utilizan en la práctica clínica, fácilmente entendidos por los pacientes y no requieren conocimientos del lenguaje biomédico.

En el mismo artículo de Moons et al., citan que los estudios pronósticos deberían incluir exclusivamente aquellos predictores que se encuentren disponibles en el tiempo en que el modelo se pretende utilizar. Si por ejemplo, el objetivo del estudio es predecir el pronóstico de un paciente en el momento de realizar el diagnóstico, los predictores que no se puedan conocer hasta que el tratamiento haya comenzado, no tienen gran valor. Por esta razón, todos los factores predictores que se han manejado en esta Tesis se han recogido en el momento en que el paciente es reclutado para el estudio.

Antes de comenzar el desarrollo de un modelo pronóstico, deben tomarse una serie de decisiones que afectarán al modelo y por tanto a las conclusiones del estudio. Entre ellas, se encuentra la de una selección clínicamente relevante de las variables que entrarán en el modelo, una evaluación de la calidad de los datos, decidir cómo manejar los datos faltantes, elegir una estrategia para seleccionar las variables relevantes en el modelo final, decidir cómo modelar las variables continuas o seleccionar los métodos que evalúen el rendimiento y la capacidad predictiva del modelo (74). Por ello, antes de la ejecución de cada uno de los modelos descritos en esta Tesis, se ha consensuado con expertos clínicos la elección de aquellas variables clínicamente relevantes que han participado en cada uno de ellos, el manejo de los datos faltantes mediante la técnica de imputación múltiple y el estudio de la bibliografía sobre los métodos más robustos que evalúen el rendimiento y capacidad predictiva de cada uno.

Una consecuencia de este proceso y del lapso de tiempo entre las publicaciones de ambos estudios, son las diferencias en el diseño y estrategia de análisis. Los criterios de inclusión son diferentes en ambos artículos. En el primero se incluyeron aquellos pacientes con una puntuación basal por encima del cambio mínimo clínicamente

relevante y aquellos que empeoraron a los tres meses de haber entrado en el estudio; mientras que en el segundo, sólo se incluyeron aquellos pacientes con una puntuación basal por encima del cambio mínimo, eliminando por tanto aquellos con una puntuación basal inferior, independientemente del resultado final que tuvieran. La razón para este cambio fue que no se consideró metodológicamente adecuado fijar como criterio de inclusión en el análisis un resultado que se obtenía 3 meses después de la visita basal.

En el primer artículo, se realizó una imputación múltiple de datos faltantes, pero los resultados mostrados son los correspondientes al análisis de casos completos, es decir, eliminando aquellos casos con observaciones faltantes. Royston et al.,(74) consideran que si menos de un 5% de las observaciones se encuentran perdidas, llevar a cabo un análisis de casos completos es sensato. La frecuencia relativa al total de la muestra de datos perdidos en el archivo de datos utilizado para este análisis, era del 4,8%. Se realizó una comparación de los resultados de los modelos imputados y los correspondientes a los datos completos, resultando prácticamente iguales. De modo que, atendiendo al principio de parsimonia, se decidió mostrar los resultados en este primer artículo correspondientes al análisis de datos completos. En el segundo artículo, el porcentaje de observaciones con al menos un dato faltante se encontraba en torno al 30%. Asumiendo un patrón de pérdida de datos aleatorio (MAR), la decisión más sensata era llevar a cabo una imputación múltiple (11,75).

Los padres de la imputación múltiple, Rubin(12) en el año 1987 y Schafer & Olsen(62) en el 1998, sugirieron que con pocos conjuntos de datos imputados (M), del orden entre 3 y 5, se obtenían excelentes resultados. El mismo Schafer, en el año 1999(63) sugiere que no son necesarios más de 10 conjuntos de datos imputados. La recomendación actual, gracias a los estudios de simulación, es llevar a cabo más imputaciones de las que se pensaban previamente como suficientes (9,38).

La Eficiencia Relativa de la imputación (RE) se define como la proporción de varianza de una estimación para un parámetro dado con M finito respecto a la varianza si M hubiera sido infinito. La Fracción de Información Perdida (FMI) se interpreta como la proporción de datos perdidos en la muestra. La Eficiencia Relativa de la imputación, la

Fracción de Información Perdida y el número de imputaciones se encuentran relacionados entre sí.^{††} La Eficiencia Relativa de la imputación en el primer artículo fue superior al 98% con 5 imputaciones y una proporción de datos faltantes de alrededor del 5%. En el segundo trabajo, esa proporción se encontraba en torno al 30%, y llevando a cabo 10 imputaciones, la Eficiencia Relativa fue del 97% (15).

Un modelo pronóstico puede desarrollarse a partir de un conjunto de variables o tener que realizar una selección previa de aquellas que resulten más relevantes. No hay un claro consenso sobre qué estrategia es mejor. Con la estrategia del modelo completo se incluyen todas las variables candidatas y no se lleva a cabo una estrategia de modelización. De esta manera se evita el sobre-ajuste, el sesgo de selección y se obtienen errores estándar y valores-p correctos (23). Sin embargo, hay situaciones en las que una gran cantidad de variables candidatas hacen imposible esta aproximación y debe realizarse una selección. Dentro de las opciones que hay, selección hacia atrás, hacia adelante o por pasos, la estrategia hacia atrás es preferible a las otras dos (74). En el primer artículo de esta Tesis, se llevó a cabo una selección de las variables candidatas mediante estrategia hacia atrás (no automática) porque no estaba claro que la presencia de todas las variables candidatas fuera relevante a la hora de pronosticar la mejoría del dolor lumbar, irradiado y discapacidad, respectivamente. En cambio, en el segundo artículo la aproximación que se llevó a cabo fue la del modelo completo. Dado que el pronóstico de algo tan complejo como la lumbalgia y la cervicalgia, está influido por condiciones que no son el propio dolor de espalda o de cuello (2), y partiendo de la gran cantidad de datos recogidos, pudo llevarse a cabo un modelo con todas las variables candidatas sin la amenaza de saturarlo.

Hay dos aspectos a considerar en la evaluación del rendimiento de los modelos pronósticos: la calibración y la discriminación. Una de las pruebas más utilizadas para evaluar la calibración en regresión logística, es la prueba de Hosmer-Lemeshow, aplicada en el primero de los artículos que forman esta Tesis Doctoral. En el segundo, se decidió sustituir esta prueba por un gráfico de calibración. El cambio hacia el segundo método fue debido a que la agrupación por deciles que se realiza para el test

^{††} En el Apéndice VI se muestra de forma detallada la fórmula correspondiente a la Eficiencia Relativa de la imputación y otros métodos para valorarla.

de Hosmer-Lemeshow puede llegar a ser muy arbitrario (6),a que la potencia es limitada para detectar una calibración pobre y es demasiado sensible en muestras muy grandes (33,74). Por tanto, se buscó para el segundo trabajo una aproximación que obtuviera un estimador no paramétrico de la relación entre las probabilidades predichas y los resultados observados mediante un *smoother*. La evaluación de la discriminación se llevó a cabo mediante el área bajo la curva ROC en el primer artículo y por medio del índice D_{xy} de Somer en el segundo, siendo ambos métodos equivalentes.

Gracias a la generación de los nomogramas, la interpretación de los modelos es sencilla, ya que puede observarse el peso relativo de cada variable sobre el pronóstico de mejora y puede ser aplicado por pacientes o profesionales sin tener que recurrir a calculadoras para conocer el pronóstico de un individuo en concreto. Con esta Tesis, se presenta una aplicación web para cada uno de los modelos pronósticos desarrollados, de modo que el interesado en conocer la probabilidad de mejoría, tan sólo debe introducir los datos correspondientes en la aplicación, y obtendrá el resultado de manera automática.

En esta Tesis, no se ha evaluado la validez interna de los modelos mediante la técnica común del *data-splitting*, porque este procedimiento tiende a dar resultados demasiado optimistas, puesto que ambas sub-muestras son muy similares (33). Son preferibles las técnicas de *bootstrapping* (6). Además, gracias a ellas se han desarrollado los gráficos de calibración.

Una de las limitaciones de esta Tesis, es que no se ha llevado a cabo una validación de los modelos en una población diferente a la que se desarrollaron. Hay evidencias en la literatura que apuntan a que en otras poblaciones, las variables que se asocian al pronóstico del dolor lumbar, son diferentes. Por ejemplo, variables psicológicas como los pensamientos catastrofistas y las creencias sobre evitación del dolor lumbar, tienen un importante impacto en la discapacidad asociada al dolor lumbar en poblaciones anglosajonas y del norte de Europa, pero nulo impacto en la española (47,71,76–79).

Otra de las limitaciones es que no se ha evaluado el impacto de aplicar estos modelos pronósticos en el cuidado diario de los pacientes. Con los resultados obtenidos en esta

Tesis, se cuantifica la contribución al pronóstico de un paciente de los antecedentes, el examen físico, las pruebas diagnósticas y tratamientos prescritos. Sería muy interesante evaluar en qué medida el poder distinguir entre pacientes con alta y baja probabilidad de mejora, afecta a las decisiones que toma el clínico a la hora de prescribir una intervención NRT, por ejemplo. Algunos estudios sugieren que el mero hecho de acceder a una regla de predicción clínica, ya tiene un efecto sobre la decisión del clínico (80), y sería muy interesante evaluarlo en un futuro.

Mostrar pruebas convincentes del impacto, ya sea positivo o negativo, de la utilización de modelos pronósticos sobre la evolución de los pacientes es algo difícil de conseguir. Steyerberg et al.,(4) publicaron en 2013 un estudio que incluía una revisión de la literatura en seis revistas biomédicas (*NEJM, JAMA, BMJ, Annals of Internal Medicine, Lancet* y *PLOS Medicine*), describiendo el número de artículos que describían el desarrollo de modelos pronósticos, la validación y la evaluación del impacto, respectivamente. Durante el periodo 2006-2009, encontraron 61 artículos que trataban sobre desarrollo, 21 describían la validación y sólo 2 la evaluación del impacto. El estudio concluye que la falta de reconocimiento como tecnología de la salud de los estudios pronósticos, puede ser una de las causas del bajo número de estudios sobre evaluación del impacto.

Uno de los cuatro cuadrantes que definen la investigación descritos por Stokes y adaptado en un artículo publicado recientemente por Chalmers et al.,(81) es el cuadrante de Richard Doll o cuadrante de la investigación aplicada pura (*pure applied research*), que viene a representar la investigación aplicada que responde a preguntas relacionadas con la práctica clínica. Esta Tesis pretende incluirse dentro de este cuadrante, y mediante la aplicación de una metodología estadística adecuada, poder responder a preguntas sobre la probabilidad de mejoría clínicamente relevante en la intensidad de dolor lumbar, irradiado a la pierna y discapacidad asociada al dolor lumbar en un registro de pacientes del Sistema Nacional de Salud por un lado; y conocer la probabilidad de mejora clínicamente relevante tras una intervención NRT en cuanto a intensidad de dolor lumbar o cervical, dolor irradiado a la pierna o a los brazos, y discapacidad asociada al dolor lumbar o cervical, respectivamente, en una cohorte de pacientes intervenidos con la NRT.

CONCLUSIONES

- El empleo de métodos estadísticos avanzados como la imputación múltiple, los modelos de regresión y métodos de re-muestreo permite conocer de forma válida qué variables se asocian a una mejoría clínicamente relevante en la intensidad de dolor lumbar, dolor irradiado a la pierna y discapacidad asociada al dolor lumbar en un registro de pacientes del Sistema Nacional de Salud a los tres meses de la primera consulta.
- El empleo de métodos estadísticos avanzados como la imputación múltiple de datos faltantes, los modelos de regresión múltiple y métodos de re-muestreo permite conocer de forma válida qué valor tienen una serie de variables en la mejora clínicamente relevante en cuanto a intensidad de dolor lumbar o cervical, dolor irradiado a la pierna o a los brazos, y discapacidad asociada al dolor lumbar o cervical, respectivamente, tras una intervención NRT en una cohorte de casi 9.000 pacientes, tres meses después de dicha intervención.
- El desarrollo de herramientas que faciliten la interpretación de los resultados del análisis estadístico, permite una mayor accesibilidad a los resultados derivados de estudios científicos.

CONCLUSIONS

- The use of advanced statistical methods as multiple imputation of missing data, multiple regression models and re-sampling methods allow to know accurately which variables are associated with a clinically relevant improvement of the intensity of low back pain, referred pain and disability in a registry of patients which belong to the National Health System in Spain, three months after their baseline visit.
- The use of advanced statistical methods as multiple imputation of missing data, multiple regression models and re-sampling methods allow to know accurately the effect size of a set of variables over the clinically relevant improvement in three different outcomes: intensity of low back or neck pain, referred pain and disability in a cohort of almost 9,000 patients who have been treated with the NRT intervention, three months after that intervention.
- Developing tools which facilitate interpretation of results in statistical analyses, allow a better accessibility to results from scientific research studies.

APÉNDICES

Apéndice I

Apéndice II

Al ejecutar la imputación múltiple y obtener m conjuntos de datos imputados sobre los que se aplican los modelos de regresión, las estimaciones resultantes se combinan para obtener una única estimación global. Sea Q un parámetro de interés (por ejemplo, un coeficiente de regresión), \hat{Q} su estimación puntual si no hubiera datos perdidos, y U la varianza estimada de \hat{Q} . Tras realizar las imputaciones y analizar los datos imputados, se han obtenido m estimaciones $\{\hat{Q}_1, \hat{Q}_2, \dots, \hat{Q}_m\}$ con sus correspondientes varianzas estimadas $\{U_1, U_2, \dots, U_m\}$. Para obtener una única estimación puntual del parámetro de interés, \bar{Q} , y su varianza asociada, T , las reglas de Rubin establecen:

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i$$

y

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B$$

donde \bar{U} representa la variabilidad intra-imputación:

$$\bar{U} = \frac{1}{m} \sum_{i=1}^m U_i$$

Y B la variabilidad inter-imputación:

$$B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2$$

Sin datos perdidos, las estimaciones \hat{Q}_i , $i = 1, 2, \dots, m$, serían idénticas, por lo que B sería igual a 0 y $T = \bar{U}$.

La construcción de los intervalos de confianza se basa en la distribución *t-Student* que sigue el estadístico $(Q - \bar{Q})/\sqrt{T}$, con v grados de libertad.(14)

$$v = (m-1)[1 + \bar{U}/(1 + m^{-1})B]^2$$

El intervalo de confianza resultante para una cantidad Q es $\bar{Q} \pm t_{v,1-\alpha/2}\sqrt{T}$ con un nivel de confianza $(1-\alpha)$.

Apéndice III

Apéndice V

CALIBRATION PLOT JOINING MULTIPLE IMPUTATION DATASETS

Royuela A, Abraira V.

BACKGROUND

Multiple imputation provides a useful strategy for dealing with data sets with missing values. Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute. These multiply imputed data sets are then analyzed by using standard procedures for complete data and combining the results from these analyses.

Validation is an essential issue in predictive regression models. Calibration and discrimination are frequently studied as performance measures, especially for binary outcomes. A calibration plot represents in an easy way the actual probability vs the predicted probability, both on average and over the whole range of predictions. Harrell⁽¹⁾ describes the plot showing the “Apparent curve” (representing predicted probability using the model against actual probability), the “Ideal curve” (representing perfect calibration) and the “Bias-corrected curve” (representing the overfitting-corrected calibration curve, obtained by Bootstrapping).

OBJECTIVES

To develop a global calibration measure that takes into account the multiple imputation procedure.

METHODS

Data from a 7,976 patients registry were used to perform a logistic regression model in order to predict spinal pain (SP) improvement. “Improvement” in SP was defined as any reduction in the corresponding score being higher than the minimal clinically important change ⁽²⁾. MCIC have been established at 30% of the baseline value, with a minimum value of 1.5 VAS (Visual Analog Scale, range from better to worse; 0-10) points for SP. Improvement in SP was the dependent variable and the independent

variables were those considered to be clinically relevant and those which had shown to have a prognostic value.

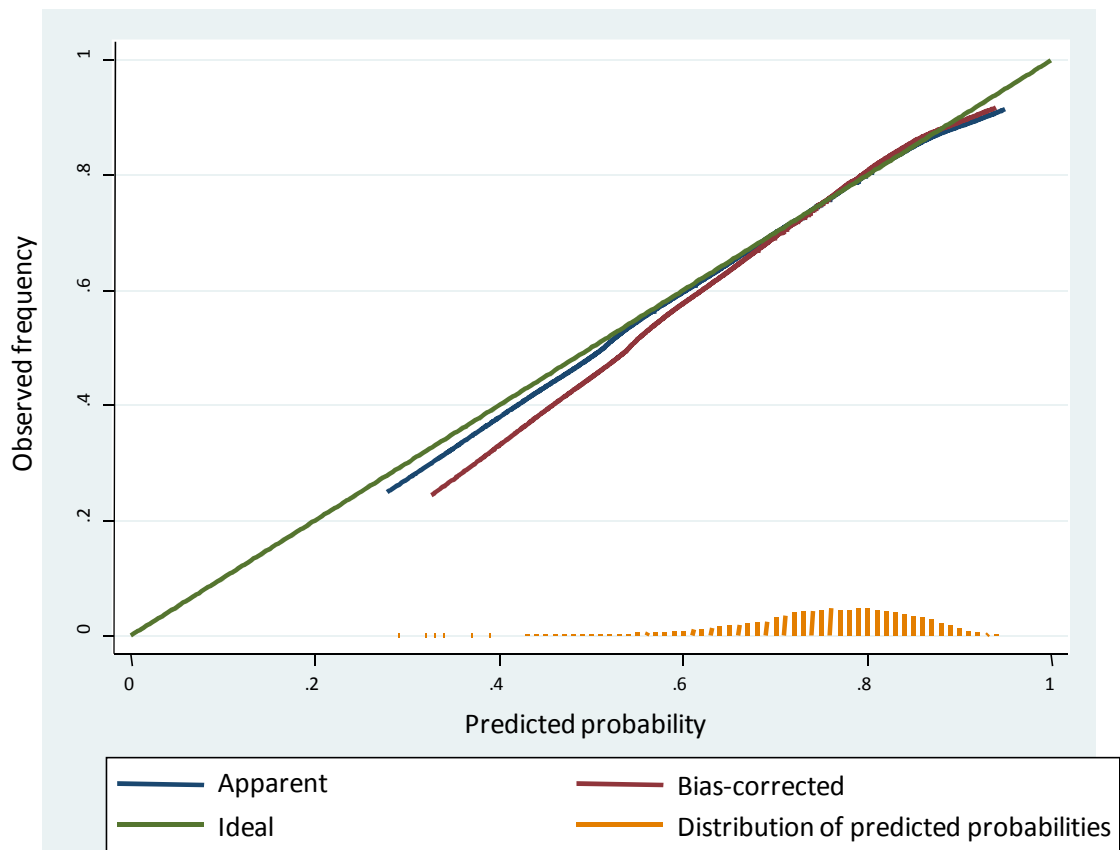
Multiple imputation by chained equations⁽³⁾ was carried out assuming a Missing At Random (MAR) pattern and 10 imputed datasets were obtained. Calibration was assessed by means of the calibration plot described by Harrell. Curves were drawn through locally weighted scatterplot smoothing. Discrimination was assessed through the D_{xy} index.

RESULTS

At discharge, a clinically relevant SP improvement was experienced by 75% of patients. Coefficients from the model were obtained in order to predict the SP improvement. Discrimination of the model was moderated (corrected D_{xy} index equal to 0.241).

A calibration plot is a plot in which observed frequencies (O) are plotted against predicted probabilities (P) ("Apparent curve"). If O and P agree over the whole range of probabilities, the plot shows a 45° line ("Ideal curve"). The intercept of the line (γ_0) is 0 and slope (γ_1) is 1. The intercept and slope of the calibration line can be estimated in a logistic regression model with the linear predictor as the only predictor variable ((log odds(improvement in SP)= $\gamma_0 + \gamma_1$ *linear predictor)).

To draw the "Bias-corrected curve", we need to join the information provided by the Multiple Imputation phase (10 datasets to assess their calibration) and the γ_0 and γ_1 estimation. In each dataset, 500 bootstrap resamples were performed in order to get the corresponding intercepts and slopes. Therefore we obtained ten intercepts and ten slopes by each time. Then averaging the 10 intercepts and 10 slopes, we achieved a global γ_0 (equal to 0.120) and a global γ_1 (equal to 0.890). All the curves were drawn through the *lowess* command in Stata (used to get smoothed means). The rug plot at the bottom of the graph indicates the distribution of predicted probabilities of spinal pain improvement.



CONCLUSION

It is feasible to draw a calibration plot which provides bias-corrected indexes and takes into account the multiple datasets obtained after a multiple imputation procedure.

REFERENCES

- (1) Harrell F. Regression Modeling Strategies with Applications to Linear Models, Logistics Regression, and Survival Analysis. New-York: Springer; 2001.
- (2) Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine (Phila Pa 1976) 2007; 32:2915-2920.
- (3) White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. StatMed 2011; 30:377-399.

CALIBRATION PLOT JOINING MULTIPLE IMPUTATION DATASETS

Royuela A^{a,b}, Abraira V^{b,a}

a) CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain ; b) Unidad de Bioestadística Clínica, Hospital Ramon y Cajal, IRYCIS, Madrid, Spain

BACKGROUND

Multiple imputation provides a useful strategy for dealing with data sets with missing values. Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute. These multiple imputed data sets are then analyzed by using standard procedures for complete data and combining the results from these analyses.

Validation is an essential issue in predictive regression models. Calibration and discrimination are frequently studied as performance measures, especially for binary outcomes. A calibration plot represents in an easy way the actual probability vs the predicted probability, both on average and over the whole range of predictions. Harrell⁽¹⁾ describes the plot showing the "Apparent curve" (representing predicted probability using the model against actual probability), the "Ideal curve" (representing perfect calibration) and the "Bias-corrected curve" (representing the overfitting-corrected calibration curve, obtained by Bootstrapping).

OBJETIVES

To develop a global calibration measure that takes into account the multiple imputation procedure.

METHODS

Data from a 7,976 patients registry were used to perform a logistic regression model in order to predict spinal pain (SP) improvement. "Improvement" in SP was defined as any reduction in the corresponding score being higher than the minimal clinically important change ⁽²⁾. MCIC have been established at 30% of the baseline value, with a minimum value of 1.5 VAS (Visual Analog Scale, range from better to worse; 0-10) points for SP.

Improvement in SP was the dependent variable and the independent variables were those considered to be clinically relevant and those which had shown to have a prognostic value.

Multiple imputation by chained equations ⁽³⁾ was carried out assuming a Missing At Random (MAR) pattern and 10 imputed datasets were obtained. Calibration was assessed by means of the calibration plot described by Harrell. Curves were drawn through locally weighted scatterplot smoothing. Discrimination was assessed through the Dxy index.

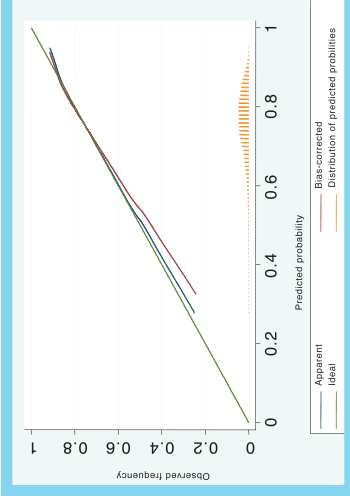
RESULTS

At discharge, a clinically relevant SP improvement was experienced by 75% of patients. Coefficients from the model were obtained in order to predict the SP improvement. Discrimination of the model was moderated (corrected Dxy index equal to 0.241).

RESULTS

A calibration plot is a plot in which observed frequencies (O) are plotted against predicted probabilities (P) ("Apparent curve"). If O and P agree over the whole range of probabilities, the plot shows a 45° line ("Ideal curve"). The intercept of the line (y0) is 0 and slope (y1) is 1. The intercept and slope of the calibration line can be estimated in a logistic regression model with the linear predictor as the only predictor variable ((log odds(improvement in SP)= y0 + y1*linear predictor)).

To draw the "Bias-corrected curve", we need to join the information provided by the Multiple Imputation phase (10 datasets to assess their calibration) and the y0 and y1 estimation. In each dataset, 500 bootstrap resamples were performed in order to get the corresponding intercepts and slopes. Therefore we obtained ten intercepts and ten slopes by each time. Then averaging the 10 intercepts and 10 slopes, we achieved a global y0 (equal to 0.120) and a global y1 (equal to 0.890). All the curves were drawn through the *lowess* command in Stata (used to get smoothed means). The rug plot at the bottom of the graph indicates the distribution of predicted probabilities of spinal pain improvement.



CONCLUSION

It is feasible to draw a calibration plot which provides bias-corrected indexes and takes into account the multiple datasets obtained after a multiple imputation procedure.

REFERENCES

(1) Harrell F. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. New-York: Springer; 2001.

(2) Kovacs FM, Abraira V, Royuela A, Corcoll J, Alegre L, Cano A et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine (Phila Pa 1976) 2007; 32:2915-2920.

(3) White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011; 30:377-399.

Apéndice VI

Entendiendo la Fracción de Información Perdida (FMI) como una medida de incertidumbre debida a la falta de respuesta; la Eficiencia Relativa (RE) como la eficiencia de la estimación para m imputaciones en relación con un número infinito de imputaciones y M el número de imputaciones, la ecuación que relaciona estos tres conceptos:

$$RE = \sqrt{\left(1 + \frac{FMI}{M}\right)}$$

En el primer artículo de esta Tesis, se llevaron a cabo 5 imputaciones, por tanto, el error estándar es tan sólo $\sqrt{\left(1 + \frac{0,05}{5}\right)} = 1,005$ veces mayor que su valor mínimo hipotético. En el segundo artículo, aplicando la misma ecuación, el error estándar para el parámetro con mayor número de datos faltantes y ejecutando 10 imputaciones es $\sqrt{\left(1 + \frac{0,3}{10}\right)} = 1,015$ veces mayor que su valor mínimo hipotético.

Por otro lado, es interesante disponer de medidas que evalúen la influencia que tienen los datos faltantes en la estimación de Q . Rubin(12) proporciona una estimación de FMI debida a los datos faltantes que viene dada para un Q escalar por medio de

$$\gamma = \frac{\left(r + \frac{2}{v+3}\right)}{r+1}$$

donde

$$r = \frac{(1 + m^{-1})B}{\bar{U}}$$

y

$$v = (m - 1) \left[1 + \frac{m\bar{U}}{(m + 1)B} \right]^2$$

El parámetro r representa el incremento relativo de la varianza debido a la no-respuesta. A partir de γ , se puede obtener de forma aproximada la RE, en términos de error estándar, mediante

$$\left(1 + \frac{\gamma}{m}\right)^{-1}$$

Cuando no hay información perdida acerca de Q , los valores de r y B son iguales a 0. Con valores altos de m o pequeños valores de r , los grados de libertad v serán elevados y la distribución se aproximará a una distribución normal.

En la siguiente tabla (Tabla 3) se muestran las eficiencias relativas (RE) para diferentes valores de m y γ . En casos en los que la proporción de información faltante es pequeña, un pequeño número de imputaciones son suficientes para el análisis de imputación múltiple.

Tabla 3. Eficiencias relativas de la imputación múltiple según diferentes valores de m y γ .

	γ				
m	10%	20%	30%	40%	50%
3	0.9677	0.9375	0.9091	0.8571	0.8108
5	0.9804	0.9615	0.9434	0.9091	0.8772
10	0.9901	0.9804	0.9709	0.9524	0.9346
20	0.9950	0.9901	0.9852	0.9756	0.9662

BIBLIOGRAFÍA

1. Calero J del R, Cabrera RH, Artalejo FR. Fundamentos de Epidemiología Clínica. Síntesis; 1996. 400 p.
2. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.
3. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338(1756-1833 (Electronic)):b375.
4. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(1549-1676 (Electronic)):e1001381.
5. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(1549-1676 (Electronic)):e1001380.
6. Harrell FE Jr, Lee KL, Mark DB. Tutorial in biostatistics: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 28 de febrero de 1996;15(0277-6715 (Print)):361-87.
7. Spratt M, Carpenter J, Sterne JA, Carlin JB, Heron J, Henderson J, et al. Strategies for multiple imputation in longitudinal studies. *AmJEpidemiol*. 15 de agosto de 2010;172(1476-6256 (Electronic)):478-87.
8. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *AmJEpidemiol*. 15 de agosto de 2008;168(1476-6256 (Electronic)):355-7.
9. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 20 de febrero de 2011;30(1097-0258 (Electronic)):377-99.
10. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338(1756-1833 (Electronic)):b2393.
11. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med*. 10 de diciembre de 2010;29(1097-0258 (Electronic)):2920-31.
12. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
13. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *IntJ Methods PsychiatrRes*. marzo de 2011;20(1049-8931 (Print)):40-9.
14. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: J. Wiley & Sons; 1987.
15. García JG, Albaladejo JP, Fernández JAM. Métodos de inferencia estadística con datos faltantes: estudio de simulación sobre los efectos en las estimaciones. *Estad Esp*. 2006;48(162):241-70.

16. van Buuren S., Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *StatMed*. 30 de marzo de 1999;18(0277-6715 (Print)):681-94.
17. Royston P, White IR. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. *J Stat Softw*. 2011;45(4):1-20.
18. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(1549-1676 (Electronic)):1-12.
19. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMCMed*. 2010;8(1741-7015 (Electronic)):21.
20. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMCMed*. 2010;8(1741-7015 (Electronic)):20.
21. Abaira V, Muriel A, Emparanza JI, Pijoan JI, Royuela A, Plana MN, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. diciembre de 2013;66(12):1340-1346.e5.
22. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *AnnInternMed*. 16 de marzo de 1999;130(0003-4819 (Print)):515-24.
23. Harrell F. *Regression Modeling Strategies with Applications to Linear Models, Logistics Regression, and Survival Analysis*. New York: Springer; 2001.
24. Burton AK, Balague F, Cardon G, Eriksen HR, Henrotin Y, Lahad A, et al. Chapter 2. European guidelines for prevention in low back pain : November 2004. *EurSpine J*. marzo de 2006;15 Suppl 2(0940-6719 (Print)):S136-S168.
25. Van Tulder MW, Tuut M, Pennick V, Bombardier C, Assendelft WJ. Quality of primary care guidelines for acute low back pain. *Spine Phila Pa* 1976. 1 de septiembre de 2004;29(1528-1159 (Electronic)):E357-E362.
26. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J*. marzo de 2006;15 Suppl 2(0940-6719 (Print)):S192-S300.
27. Grupo Español de Trabajo del Programa Europeo COST B13. Guja de Practica Clínica para la Lumbalgia Inespecífica [Internet]. 2005. Recuperado a partir de: www.REIDE.org
28. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA*. 12 de agosto de 1992;268(0098-7484 (Print)):760-5.
29. Cassidy JD, Carroll LJ, Cote P. The Saskatchewan health and back pain survey. The prevalence of low back pain and related disability in Saskatchewan adults. *Spine Phila Pa* 1976. 1 de septiembre de 1998;23(0362-2436 (Print)):1860-6.
30. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain*. enero de 2000;84(0304-3959 (Print)):95-103.

31. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J.* enero de 2008;8(1529-9430 (Print)):8-20.
32. Hepple P, Robertson AR. Back pain-- reducing long-term problems. *BrJGenPract.* mayo de 2006;56(0960-1643 (Print)):324-6.
33. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ.* 19 de agosto de 1995;311(0959-8138 (Print)):485.
34. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. *JClinEpidemiol.* diciembre de 2005;58(0895-4356 (Print)):1241-51.
35. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *JClinEpidemiol.* mayo de 2003;56(0895-4356 (Print)):395-407.
36. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine Phila Pa 1976.* 1 de diciembre de 2007;32(1528-1159 (Electronic)):2915-20.
37. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *PrevSci.* septiembre de 2007;8(1389-4986 (Print)):206-13.
38. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York: Springer-Verlag; 2002.
39. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338(1756-1833 (Electronic)):b605.
40. Jr DWH, Lemeshow S. *Applied Logistic Regression.* New York: John Wiley & Sons; 2004. 397 p.
41. Austin PC. Bootstrap model selection had similar performance for selecting authentic and noise variables compared to backward variable elimination: a simulation study. *J Clin Epidemiol.* octubre de 2008;61(1878-5921 (Electronic)):1009-17.
42. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* abril de 1982;143(0033-8419 (Print)):29-36.
43. IASP. *Classification of Chronic Pain: Description of chronic pain syndromes and definitions of pain terms.* Seattle: IASP Press; 1994.
44. Heymans MW, van BS, Knol DL, Anema JR, van MW, de Vet HC. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. *Spine J.* octubre de 2010;10(1878-1632 (Electronic)):847-56.
45. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von KM, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine Phila Pa 1976.* 1 de enero de 2008;33(1528-1159 (Electronic)):90-4.

46. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine Phila Pa* 1976. 1 de agosto de 2010;35(1528-1159 (Electronic)):E827-E835.
47. Kovacs FM, Abaira V, Zamora J, Fernandez C. The transition from acute to subacute and chronic low back pain: a study based on determinants of quality of life and prediction of chronic disability. *Spine Phila Pa* 1976. 1 de agosto de 2005;30(1528-1159 (Electronic)):1786-92.
48. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
49. StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.
50. Lenfant C. Shattuck lecture--clinical research to clinical practice--lost in translation? *N Engl J Med*. 28 de agosto de 2003;349(9):868-74.
51. Butler D. Translational research: Crossing the valley of death. *Nature*. 6 de noviembre de 2008;453(7197):840-2.
52. Kovacs FM, Abaira V, Lopez-Abente G, Pozo F. Neuro-reflexotherapy intervention in the treatment of non specified low back pain: a randomized, controlled, double-blind clinical trial. *Med Clin Barc*. 6 de noviembre de 1993;101(0025-7753 (Print)):570-5.
53. Kovacs FM, Abaira V, Pozo F, Kleinbaum DG, Beltran J, Mateo I, et al. Local and remote sustained trigger point therapy for exacerbations of chronic low back pain. A randomized, double-blind, controlled, multicenter trial. *Spine Phila Pa* 1976. 1 de abril de 1997;22(0362-2436 (Print)):786-97.
54. Kovacs FM, Llobera J, Abaira V, Lazaro P, Pozo F, Kleinbaum D. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice: a cluster randomized, controlled trial. *Spine Phila Pa* 1976. 1 de junio de 2002;27(1528-1159 (Electronic)):1149-59.
55. Urrutia G, Burton AK, Morral A, Bonfill X, Zanoli G. Neuroreflexotherapy for non-specific low-back pain. *CochraneDatabaseSystRev*. 2004;(1469-493X (Electronic)):CD003009.
56. Campillo C, Casamitjana M, Carrillo-Arias F, Royuela A, Kovacs FM, Abraria V. Post-implementation surveillance of a non-pharmacological-health-technology within a National Health Service. *Int J Technol Assess Health Care*. 2014;In press.
57. Corcoll J, Orfila J, Tobajas P, Alegre L. Implementation of neuroreflexotherapy for subacute and chronic neck and back pain within the Spanish public health system: audit results after one year. *Health Policy*. diciembre de 2006;79(0168-8510 (Print)):345-57.
58. Subdirección General de Conciertos. Especificaciones técnicas: requisitos mínimos. Intervenciones neurorreflejo terapéuticas. Madrid: Dirección General del INSALUD; 2001.
59. Grupo PINS. [Implementation of neuroreflexotherapy in the treatment of back pain. Results of a pilot trial]. *GacSanit*. julio de 2004;18(0213-9111 (Print)):275-86.

60. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomas M, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC MusculoskeletDisord*. 2008;9(1471-2474 (Electronic)):43.
61. van Buuren S. Number of publications (log) on multiple imputation during the period 1977-2010 according to three counting methods [Internet]. Multiple-imputation. 2014. Recuperado a partir de: <http://www.stefvanbuuren.nl/mi/>
62. Schafer J, Olsen M. Multiple imputation for multivariate missing data problems: a data analyst's perspective. *Multivar Behav Res*. 1998;33:545-71.
63. Schafer JL. Multiple imputation: a primer. *StatMethods MedRes*. marzo de 1999;8(0962-2802 (Print)):3-15.
64. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. enero de 2010;21(1531-5487 (Electronic)):128-38.
65. Efron B, Gong G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *Am Stat*. 1983;37(1):36-48.
66. Pallicer A, Corcoll J, Orfila J. Implementation and post-marketing surveillance of neuroreflexotherapy in the routine practice of the Spanish Public Health System. 2007. Recuperado a partir de: <http://www.lbpforum.org/Abstracts/Abstract103.jpg>
67. Huskisson EC. Measurement of pain. *Lancet*. 9 de noviembre de 1974;2(0140-6736 (Print)):1127-31.
68. Kovacs FM, Llobera J, Gil Del Real MT, Abaira V, Gestoso M, Fernandez C, et al. Validation of the spanish version of the Roland-Morris questionnaire. *Spine Phila Pa* 1976. 1 de marzo de 2002;27(1528-1159 (Electronic)):538-42.
69. Kovacs FM, Bago J, Royuela A, Seco J, Gimenez S, Muriel A, et al. Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. *BMC MusculoskeletDisord*. 2008;9(1471-2474 (Electronic)):42.
70. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine Phila Pa* 1976. 15 de septiembre de 1998;23(0362-2436 (Print)):2003-13.
71. Kovacs F, Noguera J, Abaira V, Royuela A, Cano A, Gil Del Real MT, et al. The influence of psychological factors on low back pain-related disability in community dwelling older persons. *Pain Med*. octubre de 2008;9(1526-4637 (Electronic)):871-80.
72. Barrett TW, Martin AR, Storrow AB, Jenkins CA, Harrell FE Jr, Russ S, et al. A clinical prediction model to estimate risk for 30-day adverse events in emergency department patients with symptomatic atrial fibrillation. *AnnEmergMed*. enero de 2011;57(1097-6760 (Electronic)):1-12.
73. Cook CE, Learman KE, O'Halloran BJ, Showalter CR, Kabbaz VJ, Goode AP, et al. Which prognostic factors for low back pain are generic predictors of outcome across a range of recovery domains? *Phys Ther*. enero de 2013;93(1):32-40.

74. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(1756-1833 (Electronic)):b604.
75. Janssen KJ, Donders AR, Harrell FE Jr, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol*. julio de 2010;63(1878-5921 (Electronic)):721-7.
76. Kovacs F, Abaira V, Cano A, Royuela A, Gil Del Real MT, Gestoso M, et al. Fear avoidance beliefs do not influence disability and quality of life in Spanish elderly subjects with low back pain. *Spine Phila Pa 1976*. 1 de septiembre de 2007;32(1528-1159 (Electronic)):2133-8.
77. Kovacs F, Abaira V, Santos S, Diaz E, Gestoso M, Muriel A, et al. A comparison of two short education programs for improving low back pain-related disability in the elderly: a cluster randomized controlled trial. *Spine Phila Pa 1976*. 1 de mayo de 2007;32(1528-1159 (Electronic)):1053-9.
78. Kovacs FM, Seco J, Royuela A, Corcoll-Reixach J, Pena-Arrebola A. The prognostic value of catastrophizing for predicting the clinical evolution of low back pain patients: a study in routine clinical practice within the Spanish National Health Service. *Spine J [Internet]*. 19 de julio de 2012;(1878-1632 (Electronic)). Recuperado a partir de: PM:22819318
79. Albaladejo C, Kovacs FM, Royuela A, del PR, Zamora J. The efficacy of a short education program and a short physiotherapy program for treating low back pain in primary care: a cluster randomized trial. *Spine Phila Pa 1976*. 1 de marzo de 2010;35(1528-1159 (Electronic)):483-96.
80. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. *Ann Intern Med*. 7 de febrero de 2006;144(3):201-9.
81. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. 11 de enero de 2014;383(9912):156-65.

TABLAS Y FIGURAS

Tablas

Tabla 1. Variables recogidas en la visita basal correspondientes al primer artículo de esta Tesis (pág. 72)

Tabla 2. Variables recogidas en la visita basal correspondientes al segundo artículo de esta Tesis (pág. 104)

Tabla 3. Eficiencias relativas de la imputación múltiple según diferentes valores de m y γ (pág. 190)

Figuras

Figura 1. Visualización de la aplicación web a la hora de introducir la información sobre las diferentes variables en el primer artículo de esta Tesis (pág. 78)

Figura 2. Visualización de la aplicación web mostrando la probabilidad de mejoría en los tres modelos para un caso concreto en el primer artículo de esta Tesis (pág. 79)

Figura 3. Visualización de la aplicación web a la hora de introducir la información sobre las diferentes variables en el segundo artículo de esta Tesis (pág. 110-111)

Figura 4. Visualización de la aplicación web mostrando la probabilidad de mejoría en los tres modelos para un caso concreto en el segundo artículo de esta Tesis (pág. 112)

